160-TP-006-001

Pilot Data Migration Report

Technical Paper

October 30, 1995

Technical Paper—Not intended for formal review or government approval.

Prepared Under Contract NAS5-60000

RESPONSIBLE ENGINEER

R. Suresh /s/, Jon Pals /s/, Deepti Mukund /s/ 11/3/95

R. Suresh, Team Leader Date
Jon Pals, Principal Programmer/Analyst
Deepti Mukund, Associate Programmer/Analyst
EOSDIS Core System Project

SUBMITTED BY

Joy Colucci /s/ 11/3/95

Joy Colucci, Science Office Manager Date
EOSDIS Core System Project

Hughes Applied Information Systems
Landover, Maryland

This page intentionally left blank.

Contents

1. Introduction

2. Executive Summary

3. Pilot Data Migration Process

4. Detailed Description of Tasks for Each Converted Data Product

5. Lessons Learned

6. HDF-Related Problems

7. Migration Time Estimates

Appendices

Figures

Tables

1. Introduction

1.1 Purpose

The purpose of the Pilot Data Migration Study is to better understand the problems associated with Version 0 data migration, i.e., analyze and identify problems early by migrating selected data products recommended by Earth Observing System (EOS) Data Active Archive Centers (DAAC); examine the metadata migration for selected data products; estimate time required to migrate data products; use and evaluate the Hierarchical Data Format - Earth Observing System (HDF-EOS) standards; and investigate the possibility of developing a generalized translator. This document summarizes the results of the Pilot Data Migration Study.

Because of the value of past remote sensing data to Global Change Research, NASA has elected to migrate EOSDIS Version 0 data to ECS, since ECS is being designed for a long life cycle (at least two decades beyond the launch of the first EOS spacecraft) with architectural features that facilitate technology upgrades and evolution. Migration to ECS will ensure continued maintenance and access to important data from previous NASA Earth science flight missions and other data sources. Results and “lessons learned” from the Pilot Data Migration Study are being incorporated into the Engineering Phase of the V0 Data Migration Project that is now underway by ECS.

1.2 Study Team Members

The study was performed by the following ECS team members: R. Suresh, Liping Di, Deepti Mukund, Jon Pals, and Doug Ilg (all Hughes STX), and Tom Dopplick (HAIS). The primary responsibilities of the team members were as follows: Tom Dopplick was responsible for the overall ECS technical direction; R. Suresh was the study team leader and was responsible for the implementation of the study; Deepti Mukund and Jon Pals were responsible for analysis and software development; Doug Ilg checked for compliance with ECS standards; and Liping Di participated in metadata analysis. Phil Ardanuy (Hughes STX) provided editorial comments and Ted Meyer (ESDIS) provided useful technical comments and assistance. Greg Hunolt and Bill North provided overall ESDIS direction.

1.3 Organization

This paper is organized into the following broad categories:

1. Introduction
2. Executive Summary
3. Pilot Data Migration Process
4. Detailed Description of Tasks for Each Converted Data Product
5. Lessons Learned

6. HDF-Related Problems
7. Migration Time Estimates

   Appendices A - E (Software Interface Specification)

   Appendices F - J (User Documentation)

   Acronyms

   References

1.4 Acknowledgments

We would like to thank the various people who helped with this task from the DAACs, the V0 IMS Team and the ECS Team. Special thanks go to the following people: Nettie La Belle-Hamer and Ruth Duerr (ASF DAAC), Saud Amer and Mike Neiers (EDC DAAC), A. K. Sharma, James Johnson and Daniel Ziskin (GSFC DAAC), Glenn Shirtliffe (JPL DAAC), Haldun Direskeneli and Lise Maring (LaRC DAAC), Siri Jodha Khalsa and Karen Robinson (NSIDC DAAC); Ted Johnson and Vada La Fontaine (V0 IMS Team); and Denise Heller, Janet Hylton, and Michele Kimble (ECS Team).

1.5 Review and Approval

This Technical Paper is an informal document approved at the Office Manager level. It does not require formal Government review or approval. Questions regarding technical information contained within this Paper should be addressed to either R. Suresh, 301-441-4092, suresh@ulabsgi.gsfc.nasa.gov or to Tom Dopplick, 301-925-0333, tom@eos.hitc.com.

Questions concerning distribution or control of this document should be addressed to:

Data Management Office
The ECS Project Office
Hughes Applied Information Systems
1616 McCormick Dr.
Landover, MD 20785

2. Executive Summary

Background

The purpose of the Pilot Data Migration Study was to understand the problems associated with data migration, i.e., analyze and identify problems early by migrating selected data products recommended by Earth Observing System (EOS) Data Active Archive Centers (DAAC); examine the metadata migration for selected data products; estimate time required to migrate data products; use and evaluate the Hierarchical Data Format - Earth Observing System (HDF-EOS) standards; and investigate the possibility of developing a generalized translator. This study was carried out in two phases: 1) analysis of V0 metadata and 2) migration of sample V0 granules.

1) Analysis of V0 metadata: The first phase focused on metadata migration. Detailed analysis and results of the metadata migration are described in the “ECS Core Metadata Migration Analysis, April 1995” document. The results were also presented at the EOSDIS Core System (ECS) Data Modeling Working Group workshop in May 1995.

2) Migration of sample V0 granules: The second phase focused on converting a few sample data products from their native formats to HDF-EOS. The proposed HDF-EOS standards were used wherever possible. ECS tools like the Product Generating System (PGS) Tool Kit, EOSView, etc., were examined to determine whether they could assist in the migration process. It was determined that these tools had not quite reached the stage where they could be used for this effort. The data products for migration were selected from the list of products recommended by the DAACs and are described in the Table 2-1. Products are listed alphabetically.

Table 2-1. Data Products Description.

Data Product and Science Data Plan Reference Number	Platform	Instrument	DAAC
AVHRR Global1 Km Land 10-Day Composite, E-2	NOAA - 11	AVHRR	EDC
FIRE-NWS_In_Sonde, L-55	FIRE - Field Campaign	Rawinsonde	LaRC
GPS Sea Ice Motion Vectors, A-12	ERS-1	SAR	ASF
TOPEX/Poseidon Merged Geophysical Data Record (MGDR), J-32	TOPEX /Poseidon	Altimeter	JPL
UARS SOLSTICE, G-15	UARS	SOLSTICE	GSFC

The data products listed in Table 2-1, and described in detail in later sections, vary in complexity, size, science (Land, Ocean and Atmosphere) and represent different applications.

Personnel, Hardware, and Software

Four staff members with varying experience in the use of HDF participated in this activity. This was done to determine whether a user's experience with HDF software is a factor in the amount of time required for translation; the pilot study found that HDF proficiency could be quickly acquired by team members with less HDF experience through self education and interaction with other team members who had prior HDF experience. In particular, the team member with the least HDF experience required an additional 2 weeks of self education, at the beginning of the study, to gain an understanding of the HDF library and data model. After the initial self education, the individual was able to interact effectively with the more experienced HDF developers. The prototyping work was carried out using the HDF Version 3.3 Release 4 and HDF Version 4.0 Beta 1 release of the software library on a Sun SPARC/20 workstation. The programs were written using the American National Standards Institute (ANSI) C language.

2.1 Which Data Products Should be Converted to HDF-EOS?

This prototype activity raised some very important questions. In the long term, conversion to HDF-EOS should provide improved portability, tool support and performance; increased usage by more researchers; and lower overall research costs. However, long term benefits have to be weighed against the impact on current services. For example, in the case of the field campaign data from the First ISCCP Regional Experiment (FIRE), the data product was already in ASCII. Translating the FIRE-NWS_In_Sonde data to HDF-EOS did not increase the portability of the data. In some cases, retaining the native format and migrating the metadata may be the best choice.

Should all the V0 data products be converted to HDF-EOS? The answer is no. Our experience during this exercise indicates that conversion should be done on a case-by-case basis.

2.2 Metadata Issues

Metadata attributes, as defined by the ECS core metadata Version 2 and used in this study, are too big to handle. Only about 30% of the ECS core metadata attributes are electronically available from Version 0 sources. A significant amount of work will be required to generate all 270 ECS core metadata attributes for every migrated data product. Of great importance is the need to map ECS services to the ECS core metadata attributes in order to establish the minimum required attributes for search and query, and to establish the sensitivity of ECS services to varying metadata attributes. Work is underway by ECS to perform this mapping. (Note: After completing the metadata analysis, the ECS core metadata attributes were refined and the number reduced, but the general results discussed in this document should remain valid.)

About 30% of the metadata attributes, as defined by the ECS core metadata Version 2 standard, are electronically available for the Version 0 data products that are candidates for migration to the ECS system. Many of the metadata attributes (46%) need to be provided manually. This may significantly increase the time and resources required for metadata migration.

2.3 HDF-EOS Issues

HDF-EOS was used during this study. Even though the HDF-EOS Application Program Interface (API) was not available for prototyping, HDF-EOS specification documents were used. ECS is developing prototype APIs for HDF-EOS Point, Swath, and Grid structures to be completed by February, 1996 and operational APIs to be completed by June, 1996. Writing code using a specification document may lead to multiple interpretations. This may not provide all of the expected advantages of using HDF-EOS as a standard. At this time there are no clear guidelines for choosing a particular HDF-EOS object for a particular data product. Currently, the burden lies on the implementor or user of HDF-EOS. We had problems in mapping data products into HDF-EOS objects. A clear set of guidelines from the ECS project will be helpful in implementation. Using large data products with some HDF objects creates difficulties. For example, the HDF Vdata interface is supposed to handle up to 256 fields. The Ocean Topography Experiment (TOPEX)/Poseidon Merged Geophysical Data Record (MGDR) required 124 fields. Using the current (HDF 4.0 Beta 1) Vdata interface, one could create up to 50 fields. We corrected this discrepancy by making changes to the Vdata interface. We had similar problems while using the Vshow utility. ESDIS and ECS should establish a mechanism to get any additional errors fixed within the HDF library from the National Center for Supercomputing Applications (NCSA).

Specifications of how data should be stored and organized need to be definitized for HDF-EOS Point, Swath, and Grid structures. NCSA's involvement in fixing and enhancing the HDF library will greatly help the process of data migration.

Very large data products create problems for migration. Network throughput is a critical factor. It took 16 hours to download one band (Channel 1) of the Advanced Very High Resolution Radiometer (AVHRR) Global 1 Km 10 Day Composite data product from EROS Data Center (EDC). We succeeded after several attempts. The Channel 1 data file was transferred in a non-compressed form in order to determine an approximate maximum transfer time. The Normalized Difference Vegetation Index (NDVI) data file (originally half the size of Channel 1) was transferred in a compressed form in about 1 hour. After translating the data into HDF, we could not display the data using NCSA's Collage program because of the large size. We wrote some crude subsetting programs to display this data. This illustrates the importance of subsetting and other tools for large data products.

Very large data products are inherently difficult to transfer because of network bandwidth limitations. Also, very large data products are stressing the capabilities of display and analysis tools currently available to end users. Subsetting and other tools for handling large data products are becoming increasingly important to the effective use of remotely sensed data.

2.4 Data Dictionary

A common data dictionary across DAACs and the ECS project is very important. For example, the term granule is used, defined and interpreted in many different ways. It can range from 10 Gigabytes (GB) to 1 Megabyte (MB) or less. In some cases, a granule is stored in multiple files. Hence, granule size is an important factor in migration. In general, a smaller granule size will help the user and the data producer. Available tools can generally handle small size granules. If one granule is defined as a couple of GBs, many tools will not work with this granule. Multiple file granules are a file management problem. Ideally, the granule should fit into one file.

Agreement and implementation of a common data dictionary across DAACs is important because of the widespread ambiguity in use of terms such as data set, data product, granule and file.

2.5 Migration Issues

Maintaining data organization during migration is an important issue. Generally, HDF stores data in a single file. If the data in the native format are in multiple files, migrating the data into a single HDF file may change the organization of the data. Care should be taken to ensure that the utility of the data will not be impacted due to this reorganization. For example, in the case of the Upper Atmospheric Research Satellite (UARS) Solar-Stellar Irradiance Comparison Experiment (SOLSTICE) data product, the data were organized as two files and implemented as a single HDF file. In this case, no information was lost and the organization was retained in the single HDF file. Science user input is critical to the process of organizing the data.

Good documentation of the data products will reduce the migration time. It is recommended that each data product should come with a specification document, which describes the details of the format and organization.

The file sizes before and after migration of data products vary. After converting the native formatted data into HDF, the file size was increased for two data products and decreased for three data products. The percentage changes are as follows:

AVHRR Global 1 Km Land 10 Day Composite	+ 0.03 %
FIRE-NWS_In_Sonde	+ 16.97 %
GPS Sea Ice Motion Vectors	- 36.69 %
TOPEX/Poseidon MGDR	- 22.53 %
UARS SOLSTICE	- 41.13 %

The reasons for changes in file size are explained in later sections.

The use of HDF did not significantly increase the file size of the translated data products.

2.6 Data Validation

Data validation is a very important aspect of any data migration. The highest level of confidence is needed to ensure that the migrated data are an accurate representation of the original data. This problem was approached in several ways during this study including checking the precision of numbers before and after translation, displaying and checking data visually, and in some cases by checking the contents of the data. In the case of the UARS SOLSTICE data, we identified and fixed a problem with the precision of numbers after translation. In the case of the FIRE data, we were able to compare all the values in the test data file with the output HDF file. This was possible because of the small size of the input data file.

Large data products like the AVHRR Global 1 Km 10 Day Composite are a challenge to data validation because of their size. Since the AVHRR data are images, they can be checked electronically. The image within the HDF file was an exact bit-level duplication of the original image. Non-image data are much more difficult to validate. The volume of the data product before and after migration was also checked. It was discovered that volume is not a good criteria for data validation because the volume changed after migration in all of the examples. For the operational migration of Version 0 data, a variety of tools will be used during the validation process and the results will be reviewed with the DAAC Scientist and science data provider, if available.

Data validation is a very important aspect of any data migration. The highest level of confidence is needed to ensure that the migrated data are an accurate representation of the original data.

2.7 Generalized Translator Development

We investigated the possibility of developing a generalized translator. Freeform (developed by NOAA's National Geophysical Data Center) was specifically investigated. The approach taken by many Data Description Languages, including Freeform, is to describe and store the descriptions of data. This works only for simple data products and formats such as formatted American Standard Code for Information Interchange (ASCII) text. Any data product implemented using a standard data format like HDF, network Common Data Format (netCDF), Common Data Format (CDF), or Committee on Earth Observing Satellites (CEOS) superstructure is hard to describe using this approach.

We also investigated a Commercial-Off-The-Shelf (COTS) software application called IMEX developed by Array Computing Systems. This is a generalized translator and works with simple formats like GIF, TIFF, and Sun raster. This also works with the CEOS format and the HDF Raster Image Set (RIS) object. IMEX may not work for data migration, since most of the candidate data products for migration are in other native formats. Likewise, IDL and the SFDU workbench were also examined but they do not handle an arbitrary native format.

Since many of the candidate data products for migration are in different native formats, it will be very difficult to develop a generalized translator. If there are many similar data products, it is possible to write a general translator for those products. During analysis of the Version 0 data products, a primary objective will be to identify similar products that can be migrated as a data group.

It will be very difficult to develop a generalized translator. However, if there are many similar data products, it is possible to write a general translator for those similar products.

2.8 Code Reusability

Code reusability is another technology that may reduce the code development time. In our experience, much of the code could not be reused because the five data products we studied were in different native formats. Only a small part of the code could be reused across the five products, which did not make a significant difference in time and resources. The use of Object Oriented technology would not significantly change our estimated time for the data migration task.

Code reuse is strongly dependent on the number and variations in native formats.

2.9 Configuration Control Management

Configuration control management is a very important aspect of data migration. In this pilot study, since we dealt with a small number of data products and files, configuration control was not a major issue. But, a configuration control plan is very important when dealing with a large number of data products and people. Defining standards for common naming conventions, validation schemes, test procedures, etc., will help in achieving configuration control.

An automated software tool and an audit trail are necessary for effective configuration control management during data migration of a large number of data products.

2.10 Migration Time Estimate

The time required for migration varies for different data products. This is illustrated by figures in later sections. The key factor in determining migration time is the complexity of the data product. Overall, data and metadata analysis, together with the coding and testing, comprise the largest portion of the migration effort. However, gathering documentation and other information can also require significant time. For example, it took several days to get metadata information for the AVHRR Global 1 Km Land 10 Day Composite data because the metadata are stored in a separate database. For complex data formats, understanding the input data can also require significant time.

Based on our experience in this study, it will take 2 to 3 months to develop and test prototype software, or 3 to 4 months to develop and test operational software, in order to convert one V0 data product to HDF-EOS.

2.11 Advantages of Migration to HDF-EOS

What did we gain by migrating data products into HDF-EOS? In almost all cases, it increased the portability of the data, except in the case of the FIRE-NWS_In_Sonde field campaign data product which was already in ASCII. The HDF implementation works with NCSA tools (Mosaic, Collage, Image, Xdata slice), other public domain tools (EOSView, Geoview) and commercial tools (Spyglass, Interactive Data Language (IDL)). In some cases, efficiency of the data product improved in terms of size. The HDF files were smaller in three cases: GPS Sea Ice Motion Vectors, TOPEX Poseidon MGDR, and UARS SOLSTICE. For two cases, i.e., the AVHRR Global 1Km 10 Day Composite and the FIRE-NWS_In_Sonde data, the HDF files were bigger than the native file. Especially important are the improved data-type services being developed by ECS, such as subsetting and subsampling, that will only be available for data products in HDF-EOS.

Usability testing of the converted HDF-EOS products by science end users was beyond the scope of this study. However, ECS is investigating possible approaches for initiating follow-on usability testing through ECS sponsored activities such as the ECS Collaborative Prototyping Program.

The prototype software developed for this task has not been formally reviewed by the DAAC's, but is available for testing purposes. This software can be obtained by contacting Tom Dopplick (ECS), 301-925-0333, tom@eos.hitc.com.

By converting data products to HDF-EOS, generalized data-types services, such as subsetting and subsampling, will be provided by ECS across a broad range of current and future data products.

This page intentionally left blank.

3. Pilot Data Migration Process

The purpose of the Pilot Data Migration Study was to determine the resources and cost involved in migrating a variety of EOSDIS Version 0 data products and metadata to ECS standards. The resources and cost involved were determined by examining the following aspects of data migration:

Metadata Migration Process

The metadata migration process consisted of analyzing the metadata for 34 data products and determining the availability of attributes for populating the ECS Core Metadata database. Detailed analysis and the results of the metadata analysis were described in the “ECS Core Metadata Migration Analysis, April, 1995” document. The results were also presented at the ECS Data Modeling Working Group workshop in May, 1995. About 30% of the metadata attributes, as defined by the ECS core metadata Version 2 standard, are electronically available for the Version 0 data products that are candidates for migration to ECS. Many of the metadata attributes (46%) need to be provided manually. About 20% of the attributes are not available or applicable for the data products analyzed. About 4% of the attributes were ambiguous and not categorized.

Data Migration Process

The data migration process consisted of gathering five sample V0 data products and information describing the data products from various DAACs, and converting the data products to the proposed HDF-EOS standards. The six tasks involved in the data migration process, after the data product selection, were: 1) information gathering, 2) data acquisition, 3) data and metadata analysis, 4) HDF-EOS object selection, 5) writing documentation, and 6) coding and testing. The problems uncovered during the data migration were documented as “lessons learned”.

Evaluation of Existing Tools

This evaluation investigates the possibility of using existing tools to develop a generalized translator. Freeform, developed by NOAA’s National Geophysical Data Center, was specifically investigated. The approach taken by Data Description Languages, including Freeform, is to describe the data and store the descriptions of data. This works only for simple data products and formats such as the ASCII format. Any data product implemented using a standard data format like HDF, netCDF, CDF, and CEOS is hard to describe using this approach. We also investigated a COTS software application called IMEX developed by Array Computing Systems. This is a generalized translator and works with simple formats like GIF, TIFF, and Sun raster, as well as with two more complex formats, the CEOS and the HDF RIS object formats. IMEX may not work for data migration, since most of the candidate data products for migration are in native formats. Likewise, IDL and the SFDU workbench were also examined but they do not handle an arbitrary native format. Since most of the candidate data products for migration are in different native formats, it will be very difficult to develop a generalized translator. If there are many similar data products, it is possible to write a general translator for those products.

3.1 Description of Data Products Chosen for Pilot Data Migration

The data products chosen for conversion to HDF-EOS were selected from a list of preferred data products supplied by the DAACs. An effort was made to select a data product from each DAAC for conversion. Time constraints allowed only five data products to be converted. The data products selected vary in terms of size and complexity as shown in Table 3-1, which summarizes the data product sizes, organization and the HDF-EOS objects used to represent them.

Table 3-1. Data Product Organization.

Data Product and SDP Reference Number	Organization of Input Data	Total Granule Size	Organization of Output Data
AVHRR Global 1 Km Land 10 Day Composite, E-2	1 metadata file + 1 data file	10 GB (10 files) *****	Grid
FIRE-NWS_In_Sonde, L-55	1 file with data + metadata	3.3 MB (67 files)	Point
GPS Sea Ice Motion Vectors, A-12	2 metadata files + 1 data file	28 KB (3 files)	Vdata
TOPEX/Poseidon Merged Geophysical Data Record (MGDR), J-32	1 file with data + metadata	1.1 MB (1 file)	Swath
UARS SOLSTICE, G-15	1 metadata file + 1 data file	13 KB (2 files)	Vdatas

***** Note: Only one of the 10 AVHRR files was actually migrated to HDF.

The following five data products were converted from their native format to HDF-EOS:

1. The AVHRR Global 1 Km Land 10 Day Composite data product consists of images taken of land and coastal zones in visible, near-infrared, and thermal spectrums. The AVHRR instrument's platform is the NOAA-11 satellite. The AVHRR data are archived at the EROS Data Center (EDC) DAAC.

   The data contain the following 10 bands: channel 1 (visible spectrum), channel 2 (near-infrared spectrum), channels 3 through 5 (thermal spectrums), NDVI, satellite zenith, solar zenith, relative azimuth, and date index.

The first 5 bands each have a size of approximately 1,324 MB. The last 5 bands each have a size of approximately 662 MB. The total size for a granule is approximately 10 GB. Due to the large volume of the data, only the NDVI image was converted to HDF-EOS for this data product. The AVHRR Global 1 Km Land 10 Day Composite data were represented in HDF using the HDF-EOS Grid data object.

2. The FIRE Cirrus II National Weather Service Inner-network Rawinsonde (NWS_IN_SONDE) data consist of Rawinsonde readings collected at various ground stations. The FIRE-NWS_In_Sonde data are archived at the Langley Research Center (LaRC) DAAC.

   The data contain values for 21 different variables such as pressure, temperature, relative humidity, dew-point temperature, wind speed, and wind direction.

   Each data file is approximately 50,000 bytes long. Since a granule consists of all data files recorded at a particular ground station, the granule size is variable. The FIRE-NWS_In_Sonde data were represented in HDF using the HDF-EOS Point data object.

3. The Geophysical Processor System (GPS) Sea Ice Motion Vectors data represent how far ice features moved and through what angle they rotated between two successive Synthetic Aperture Radar (SAR) images. The SAR instrument's platform is the European Space Agency's (ESA) first European Remote Sensing satellite (ERS-1). The GPS Sea Ice Motion Vectors data are archived at the Alaska SAR Facility (ASF) DAAC.

   The data cover approximately a 100 kilometer x 100 kilometer area with ice features measured within 5 kilometer cells. The data contain an ice feature's initial latitude, initial longitude, final latitude, final longitude, X displacement in kilometers, Y displacement in kilometers, rotation angle, and a reliability measure. The data are derived from geocoded low resolution SAR imagery.

   This product is generally organized as five separate files when recorded on tape in the CEOS format. When distributed as disk files, the ASF DAAC converts the product into 3 separate files: a leader file, a data file, and a trailer file. The leader and trailer files are each approximately 1,300 bytes long. The data file is approximately 25,000 bytes long. The GPS Sea Ice Motion Vectors data were represented in HDF using the HDF Vdata object.

4. The TOPEX/Poseidon Merged Geophysical Data Record (MGDR) product consists of radar altimetry data measuring sea surface level. The platform is the TOPEX/Poseidon satellite. The data are archived at the Jet Propulsion Laboratory (JPL) DAAC.

   The data contain 124 different data parameters such as altimeter ranges of sea level, wave height, wind speed, brightness temperatures, height of tides, and sensor conditions.

   Each data product consists of a cycle header file and a maximum of 254 pass files per cycle. A single pass file is considered to be a granule and is approximately 1,100,000 bytes long. The TOPEX/Poseidon MGDR data were represented in HDF using the HDF-EOS Swath Data object.

5. The UARS SOLSTICE data consist of solar spectral irradiance measurements. The data are archived at the Goddard Space Flight Center (GSFC) DAAC.

   The data contain readings representing a daily mean solar spectrum between 115.5 and 420.5 nanometers. Also included are daily average integrated intensities for selected solar spectral features.

   Each data product granule consists of a metadata file and a data file. The metadata file is approximately 700 bytes long. The data file is approximately 12,500 bytes long. The UARS SOLSTICE data were represented in HDF using the HDF Vdata object.

3.2 Data Migration Tasks Descriptions

The data migration methodology consisted of the following 6 migration tasks plus time estimates:

3.2.1 Information Gathering

This task consisted of acquiring the information that described the native format of the data products and their associated metadata. The information describing the data products was accessed using the World Wide Web (WWW), received with the sample data, or was supplied by the ECS DAAC Science Liaisons.

3.2.2 Data Acquisition

This task consisted of acquiring sample data granules for each of the data products. The sample data were retrieved from the DAACs using File Transfer Protocol (FTP), WWW or a CD-ROM.

3.2.3 Data and Metadata Analysis

This task involved studying available data and metadata information such as the description of the native format of the data products. The DAAC-supplied read software was built and run to assist with examining the data contents.

3.2.4 HDF-EOS Object Selection

This task consisted of choosing the HDF-EOS object which could best represent the data and metadata of the sample data granule. The HDF and HDF-EOS documentation were examined to determine the HDF-EOS objects which would be most appropriate for that particular data product.

3.2.5 Writing Documents

A Software Interface Specification (SIS) was written for each of the data products that describes the design chosen for the pilot migration (See Appendices A-E). Diagrams and implementation details are specified in the Appendices.

User Documentation was written to describe the compilation and use of the conversion software programs for each migrated data product (See Appendices F-J). This documentation is similar to README files supplied by the DAACs with the data product read software.

3.2.6 Coding and Testing

Conversion programs were developed to implement the design described in each SIS. The programs were written in ANSI C on a Sun SPARC/20 workstation. Testing of the generated HDF-EOS files was performed using NCSA's Mosaic and Vshow programs. The HDF data file values displayed by these utilities were compared with the original data values to check the validity of the HDF file contents. Two types of data validation were used: file size examination and data value examination.

3.2.7 Migration Time Estimates

The amount of time required for each of the six tasks was estimated by the team member performing the task and the estimates were summarized by data product and across all data products.

3.3 Migration Time Estimates Across All Data Products

Figure 3-1 summarizes the percentage of time taken to perform each of the six data migration tasks: information gathering, data acquisition, data and metadata analysis, HDF-EOS object selection, writing documents, and coding and testing.

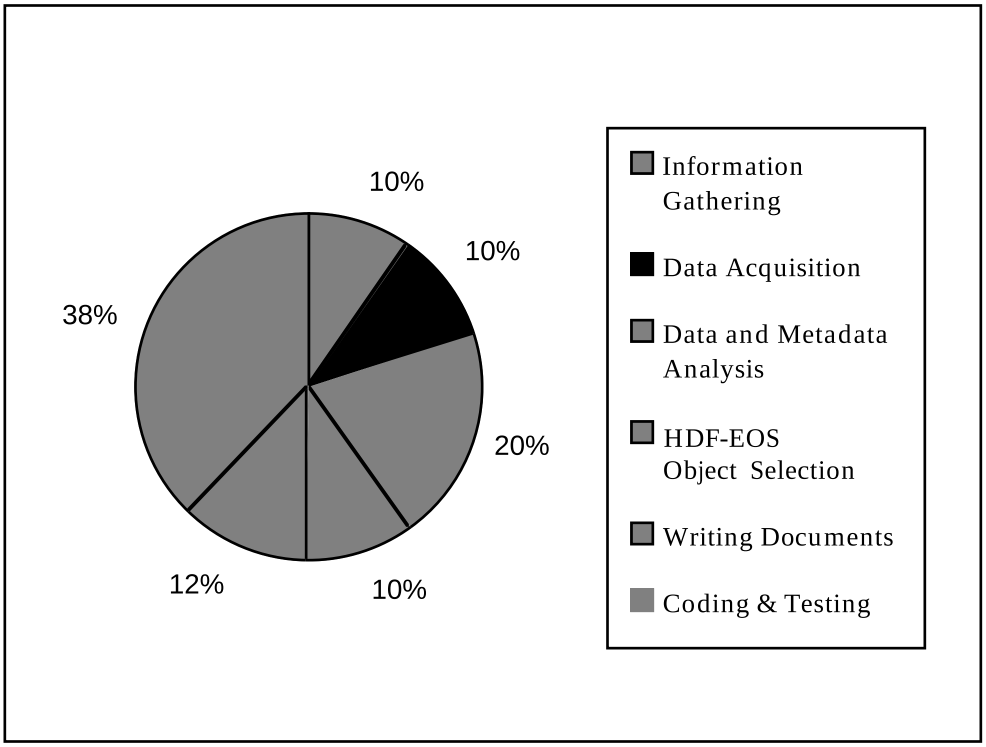

Figure 3-1. Migration Tasks Time Distribution.

The percentages in the above pie chart were calculated by computing the task percentages for each of the five data products individually. The results were then combined such that the time taken for each migrated data product was weighted evenly in the final results.

The division of the various tasks in the chart is somewhat arbitrary; tasks are not strictly sequential in all cases as there were overlapping activities between some tasks. However, we found the data and metadata analysis task together with the coding and testing task comprise the largest portion of the migration effort. All other tasks are smaller and approximately equal. Because of the limited number of products converted in this study, these results must be considered as rough estimates.

We have not gone through a formal validation process involving the DAACs since this study is a prototype activity. Additional time will be required in actual operations for DAAC validation (see Section 7.2 for additional discussion). Note: the time unit "person day" used in the document is a work day of one individual; 5 person days = 1 work week = 1 calendar week.

4.1 Detailed Description of Tasks for Each Converted Data Product

4.1 AVHRR Global 1 Km Land 10 Day Composite

4.1.1 Information Gathering

This task consisted of collecting the information that describes the native format of the AVHRR Global 1 Km Land 10 Day Composite data and metadata. The data format information was downloaded from the EDC DAAC's World Wide Web (WWW) home page. The metadata information was received as e-mails. It took several phone calls and e-mails to locate all of the available information. It took about 2 person days to get the data format information. The metadata information was gathered over a period of 8 person days time.

4.1.2 Data Acquisition

This task consisted of collecting a part of the AVHRR Global 1 Km Land 10 Day Composite granule for conversion to HDF. Due to the large size of a granule, only two bands (Channel 1 and the Normalized Difference Vegetation Index) of the data were acquired. There was no read software available. But, since the data are images, they can be displayed (after subsetting) with any image processing software. The data files were downloaded from EDC's Web page. The ECS DAAC Liaison Scientist at the EDC DAAC was called to get the procedure for downloading the data. The time to acquire the compressed NDVI data file was about two hours (one hour to transfer + one hour to uncompress). The time to acquire the Channel 1 data file (not compressed) was about 16 hours. The Channel 1 data file was acquired in order to determine an approximate maximum transfer time for a band of data. The Channel 1 data file was not converted to HDF. It took about 1 person day to get the download information. It took about 4 person days (using e-mail) to verify that the data that were downloaded were identical to what were sent.

4.1.3 Data and Metadata Analysis

This task consisted of reading the format description documents and understanding the data representation in the data granule. The metadata is stored in a separate database and is not part of the data granule. An HSTX-developed program was used to display a hexadecimal and ASCII representation of the data to see exactly what the data file contained. The documents were analyzed in about 3 person days.

4.1.4 HDF-EOS Object Selection

This task consisted of choosing which HDF-EOS objects would be best to represent the various data and metadata in an AVHRR granule. The HDF documentation and the draft HDF-EOS Primer were examined to determine which HDF objects were appropriate. This took about 2 person days.

4.1.5 Writing Documents

This task consisted of writing a preliminary Software Interface Specification (SIS) describing the input data format and the output HDF file (See Appendix A). Some implementation details such as the names and classes of objects were firmed up during the writing. User Documentation was written to describe the compilation and use of the conversion program (See Appendix F). The writing took approximately 2.5 person days.

4.1.6 Coding and Testing

This task consisted of implementing the design described in the SIS and testing the implementation. The HDF conversion program was written in ANSI C on a Sun SPARC/20. Since the metadata was not electronically-readable, a file to contain the metadata had to be created with an editor in order to test the program. The testing was done using the National Center for Supercomputing Applications' (NCSA) Mosaic and Vshow programs and UNIX's cmp program. An HSTX-developed program was written to help compare the data in the HDF file with the original data to ensure that the values had not changed. After the HDF file is created, the data within the HDF file must be validated to ensure that no data values have been lost and that no data values have changed or lost precision. The two types of data validation used were file size examination and data value examination. The coding and testing took approximately 8 person days.

4.1.6.1 File Size Examination

File size examination consisted of comparing the size of the HDF file to the size of the input NDVI file (see Table 4-1). If the HDF file size is significantly different from the input file sizes, the difference must be explained. In this case, the HDF file is approximately 235,000 bytes larger than the native NDVI file. The size difference is due to the addition of the metadata and some HDF overhead.

Table 4-1. AVHRR Global 1 Km Land 10 Day Composite Data File Size vs HDF File Size Comparison.

Data + Metadata File Size	Output HDF File Size	Size Difference	% Difference
694,418,009 bytes	694,653,395 bytes	+235,386 bytes	+0.034%

4.1.6.2 Data Value Examination

Data value examination consisted of examining the metadata and data values with the NCSA Mosaic program and Vshow utility, and the UNIX operating system's cmp program. The Mosaic and Vshow programs were used to examine the metadata attributes and the structure of the Vgroup. They were also used to examine the name, size, class and field names of the 'Geometry' Vdata.

Due to the large size of the NDVI data, manual inspection of the data values is not feasible. A program was written to extract the NDVI data from the HDF file and write it to a separate file. The new file was compared to the original NDVI data file with the UNIX program cmp.

4.1.7 Migration Time Estimates

Figure 4-1 summarizes the amount of time taken for each of the tasks described above. The total number of person days represented here is the total time taken for the various tasks.

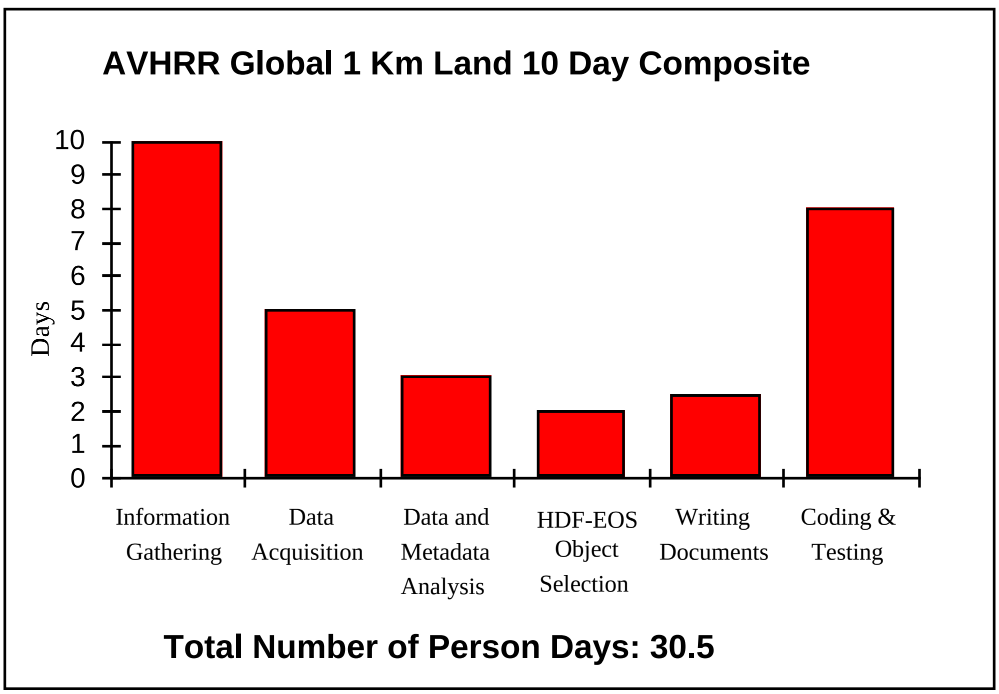

Figure 4-1. Time Taken for the Migration Tasks: AVHRR Global 1 Km Land 10 Day Composite Data.

4.2 FIRE-NWS_In_Sonde

4.2.1 Information Gathering

This task consisted of collecting the information that describes the native format of the FIRE Cirrus II NWS_IN_SONDE data. The data and metadata information was received as a readme file along with the data. This task took approximately 1 person day.

4.2.2 Data Acquisition

This task consisted of collecting a granule of the FIRE-NWS_In_Sonde data to be converted to HDF. A granule of the FIRE-NWS_In_Sonde data was ordered using the LaRC DAAC's local IMS which sent information about the account and password for an FTP site from where the data could be downloaded. The data were easily accessible and were acquired very quickly. The data came with read software and a readme file. This task was completed in 1 person day.

4.2.3 Data and Metadata Analysis

This task consisted of 1) reading the metadata and data description documents and 2) building and running the data read software.

1) Reading metadata and data description documents. This consisted of reading the format description documents and understanding the metadata and data representations in the data granule. The documents were analyzed in about 2.5 person days.

2) Building and running the data read software. This consisted of compiling and linking the given read software source code and then running the created software on the sample data granules. The output from the read software was compared with the documented data format to help understand how the data were represented in the data granule. There was a problem with building the read software. The ECS DAAC Science Liaison at the LaRC DAAC was contacted for help in this matter and the read software was built, with help from the LaRC DAAC, in 2 person days.

4.2.4 HDF-EOS Object Selection

This task consisted of choosing which HDF-EOS objects would be best to represent the various data and metadata in a FIRE-NWS_In_Sonde granule. The HDF documentation and the draft HDF-EOS Primer were examined to determine which HDF-EOS objects were appropriate. This took about 2 person days.

4.2.5 Writing Documents

This task consisted of writing a preliminary Software Interface Specification (SIS) describing the input data format and the output HDF file (See Appendix B). Some implementation details such as the names and classes of objects were firmed up during the writing. User Documentation was written to describe the compilation and use of the conversion program (See Appendix G). The writing took approximately 2.5 person days.

4.2.6 Coding and Testing

This task consisted of implementing the design described in the SIS and testing the implementation. The HDF conversion program was written in ANSI C on a Sun SPARC/20. The data values in the FIRE-NWS_In_Sonde file have been formatted in ASCII, so the data values were converted to binary values and then written out to the output HDF file. The testing was done using NCSA's Mosaic and Vshow programs. The FIRE-NWS_In_Sonde data have 22 parameters and Vshow was unable to display all 22 of them. The Vshow code was then analyzed and corrected so that all 22 parameters could be read by Vshow. The original data values displayed by the data read software were then compared with the HDF file data values displayed by Vshow to ensure that the values had not changed. The coding and testing took approximately 11.5 person days. After the HDF file is created, the data within the HDF file must be validated to ensure that no data values have been lost and that no data values have changed or lost precision. Two types of data validation were used: file size examination and data value examination.

4.2.6.1 File Size Examination

File size examination consisted of comparing the size of the HDF file to the size of the input FIRE-NWS_In_Sonde file (see Table 4-2). If the HDF file size is significantly different from the input file size, the difference must be explained. In this case, the HDF file is 8,709 bytes larger than the FIRE-NWS_In_Sonde file. The size difference is due to the addition of the metadata and some HDF overhead.

Table 4-2. FIRE-NWS_In_Sonde Data File Size vs HDF File Size Comparison.

Data File Size	Output HDF File Size	Size Difference	% Difference
51,325 bytes	60,034 bytes	+8,709 bytes	+16.97%

4.2.6.2 Data Value Examination

Data value examination consisted of examining the metadata and data values with the NCSA Mosaic program and Vshow utility. The Mosaic and Vshow programs were used to examine the metadata attributes and the structure of the Vgroup. They were also used to examine the name, size, class and field names of the 'PointIndex' and the ‘PointList’ Vdatas. The Vshow utility was used to examine the actual data values within the data Vdata. The examined data values were compared with the data values in the FIRE-NWS_In_Sonde read software output.

4.2.7 Migration Time Estimates

Figure 4-2 summarizes the amount of time taken for each of the tasks described above. The total number of person days represented here is the total time taken for the various tasks.

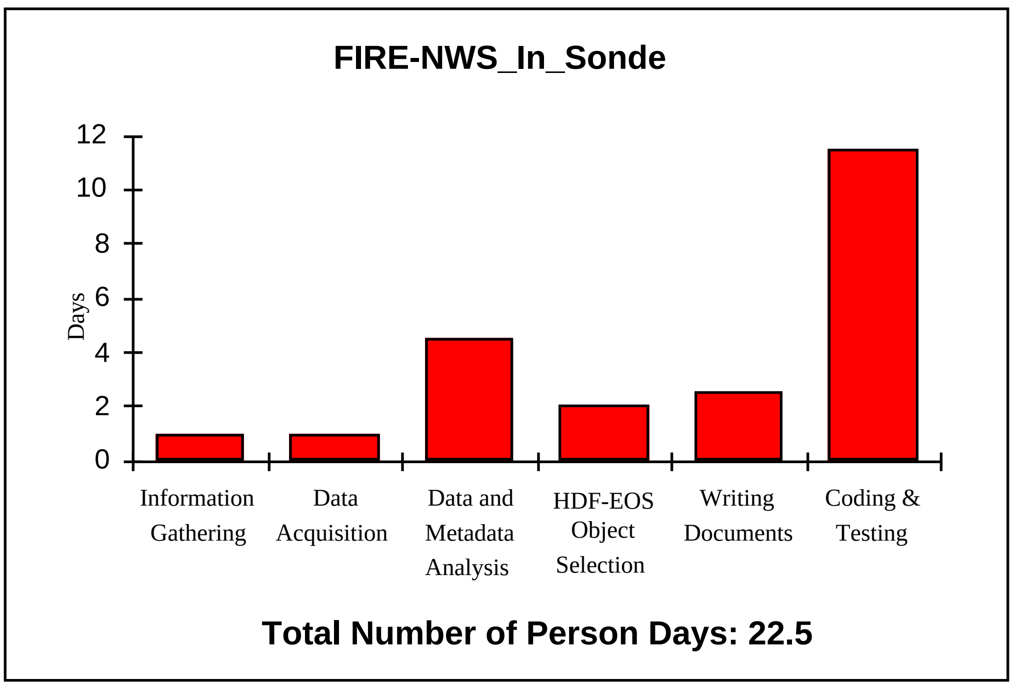

Figure 4-2. Time Taken for the Migration Tasks: FIRE-NWS_In_Sonde Data.

4.3 GPS Sea Ice Motion Vectors

4.3.1 Information Gathering

This task consisted of collecting the information that describes the native format of the GPS Sea Ice Motion Vectors data and metadata. This information was downloaded from the ASF DAAC's World Wide Web (WWW) page. An e-mail was sent to the ECS DAAC Science Liaison at the ASF DAAC to locate the information. It took about 1 person day to get the on-line data format information; whereas, it took approximately 6 person days to get the off-line information. Since we ultimately used the on-line information, we used 1 person day as the time for information gathering.

4.3.2 Data Acquisition

This task consisted of collecting a granule of data to be converted to HDF. The data were acquired from the ECS DAAC Science Liaison’s anonymous FTP site. It took about 3 person days to get the data staged on the FTP site and about 30 minutes to download the data. There does not appear to be any readily-accessible read software available for the disk data files.

4.3.3 Data and Metadata Analysis

This task consisted of reading the format description documents and understanding the data representation in the data granule. An HSTX-developed program was used to display a hexadecimal and ASCII representation of the data to see exactly what the data file contained. The documents were analyzed in about 5 person days because two data products (Sea Ice Motion Vectors and Sea Ice Type Classification) were being examined.

4.3.4 HDF-EOS Object Selection

This task consisted of choosing which HDF-EOS objects would be best to represent the various data and metadata in a GPS Sea Ice Motion Vectors granule. The HDF documentation and the draft HDF-EOS Primer were examined to determine which HDF-EOS objects were appropriate. This task took about 1 person day.

4.3.5 Writing Documents

This task consisted of writing a preliminary Software Interface Specification (SIS) describing the input data format and the output HDF-EOS file (See Appendix C). Some implementation details such as the names and classes of objects were firmed up during the writing. User Documentation was written to describe the compilation and use of the conversion program (See Appendix H). The writing took approximately 0.5 person day.

4.3.6 Coding and Testing

This task consisted of implementing the design described in the SIS and testing the implementation. The HDF-EOS conversion program was written in ANSI C on a Sun SPARC/20. The testing was done using NCSA's Mosaic and Vshow programs. An HSTX-developed program was written to help generate a printout of the original data for comparing with the Vshow output. The HDF file data values, displayed by Vshow, were compared with the original data to ensure that the values had not changed. The coding and testing took approximately 3 person days. After the HDF file is created, the data within the HDF file must be validated to ensure that no data values have been lost and that no data values have changed or lost precision. Two types of data validation were used: file size examination and data value examination.

4.3.6.1 File Size Examination

File size examination consisted of comparing the size of the HDF file to the size of the input GPS Sea Ice Motion Vectors file (see Table 4-3). If the HDF file size is significantly different from the input file sizes, the difference must be explained. In this case, the HDF file is over 9,000 bytes smaller than the GPS Sea Ice Motion Vectors leader, data and trailer files. The size difference is due to the following:

a. The data values were converted from ASCII representation to binary representation which occupies less space.

b. The trailer file (1,296 bytes long) is an almost exact duplicate of the header file and was not used at all.

c. The CEOS format contains information which was not stored in the HDF file (12 bytes per record + 720 bytes per file).

Table 4-3. GPS Sea Ice Motion Vectors Data File Size vs HDF File Size Comparison.

Data + Metadata File Size	Output HDF File Size	Size Difference	% Difference
26,883 bytes	17,019 bytes	-9,864 bytes	-36.69%

4.3.6.2 Data Value Examination

Data value examination consisted of examining the metadata and data values with the NCSA Mosaic program and Vshow utility. The Mosaic and Vshow programs were used to examine the metadata attributes. They were also used to examine the name, size, class and field names of the Data Vdata. The Vshow utility was used to examine the actual data values within the Data Vdata. The examined data values were compared against the data values extracted from the original data file by an HSTX-developed program.

4.3.7 Migration Time Estimates

Figure 4-3 summarizes the amount of time taken for each of the tasks described above. The total number of person days represented here is the total time taken for the various tasks.

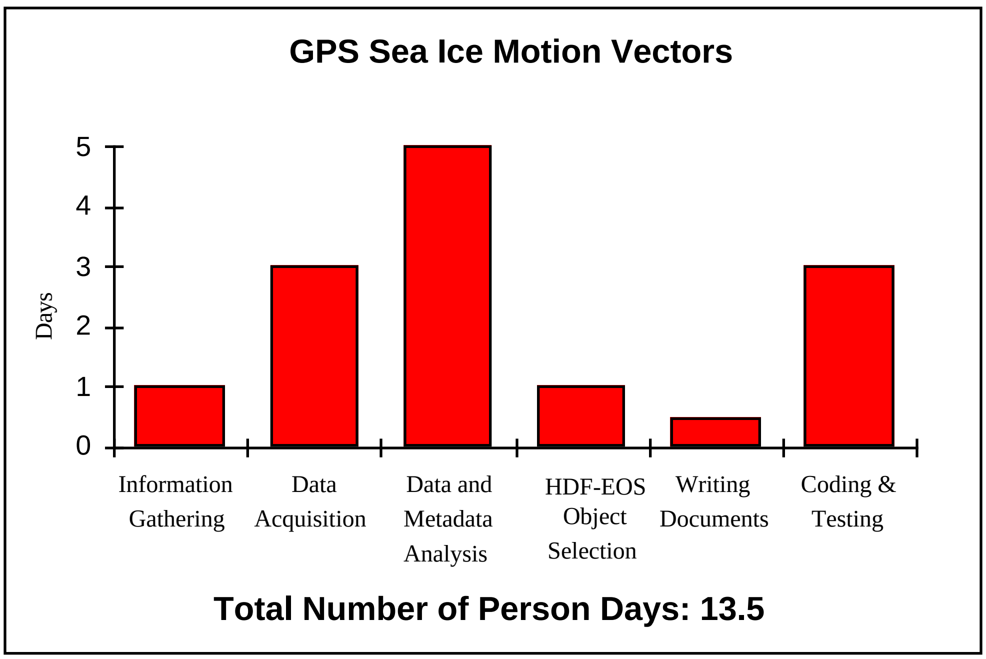

Figure 4-3. Time Taken for the Migration Tasks: GPS Sea Ice Motion Vectors Data.

4.4 TOPEX/Poseidon Merged Geophysical Data Record (MGDR)

4.4.1 Information Gathering

This task consisted of collecting the information that describes the native format of TOPEX/Poseidon MGDR data. The data and metadata information was received as a readme file along with the data. This task was completed in 1 person day.

4.4.2 Data Acquisition

This task consisted of collecting a granule of the TOPEX/Poseidon MGDR data which are distributed by the JPL DAAC on CD-ROM's. The CD-ROM was easily acquired and the data were accessed very quickly. The data came with read software and a readme file. This task was completed in 0.5 person day.

4.4.3 Data and Metadata Analysis

This task consisted of 1) reading the metadata and data description documents and 2) building and running the data read software.

1) Reading metadata and data description documents**.** This task consisted of reading the format description documents and understanding the metadata and data representations in the data granule. The documents were analyzed in about 5 person days.

2) Building and running of data read software**.** This task consisted of compiling and linking the given read software source code and then running the created software on the sample data granules. The output from the read software was compared with the documented data format to help understand how the data were represented in the data granule. The read software was built in 1 person day.

4.4.4 HDF-EOS Object Selection

This task consisted of choosing which HDF-EOS objects would be best to represent the various data and metadata in a TOPEX/Poseidon MGDR granule. The HDF documentation and the draft HDF-EOS Primer were examined to determine which HDF-EOS objects were appropriate. This task took a long time since a granule for the TOPEX/Poseidon MGDR data was very difficult to decipher. To help in this process, a scientific user for this data product was consulted. It was determined that a single 'pass file' of a 'cycle' could be considered as a granule of this data for conversion to HDF. This task took about 5 person days.

4.4.5 Writing Documents

This task consisted of writing a preliminary Software Interface Specification (SIS) describing the input data format and the output HDF-EOS file (See Appendix D). Some implementation details such as the names and classes of objects were firmed up during the writing. User Documentation was written to describe the compilation and use of the conversion program (See Appendix I). The writing took approximately 2.5 person days.

4.4.6 Coding and Testing

This task consisted of implementing the design described in the SIS and testing the implementation. The HDF-EOS conversion program was written in ANSI C on a Sun SPARC/20. The testing was done using NCSA's Mosaic and Vshow programs. The TOPEX/Poseidon MGDR data have 124 parameters; however, neither HDF4.0b1 nor Vshow (of the HDF4.0b1 version) were able to handle 124 parameters. The HDF4.0b1 and the Vshow code was analyzed and corrected so that all the 124 parameters could be handled. NCSA was also notified about this problem. The original data values displayed by the data read software were then compared with the HDF file data values displayed by Vshow to ensure that the values had not changed. The coding and testing took approximately 12 person days. After the HDF file is created, the data within the HDF file must be validated to ensure that no data values have been lost and that no data values have changed or lost precision. Two types of data validation used: file size examination and data value examination.

4.4.6.1 File Size Examination

File size examination consisted of comparing the size of the HDF file to the size of the input TOPEX/Poseidon MGDR file (see Table 4-4). If the HDF file size is significantly different from the input file sizes, the difference must be explained. In this case, the HDF file is 219,803 bytes smaller than the TOPEX/Poseidon MGDR file. The size difference is due to the deletion of spare or fill values from the input file which the HDF file does not contain.

Table 4-4. TOPEX /Poseidon MGDR File Size vs HDF File Size Comparison

Data + Metadata File Size	Output HDF File Size	Size Difference	% Difference
975,680 bytes	755,877 bytes	-219,803 bytes	-22.53%

4.4.6.2 Data Value Examination

Data value examination consisted of examining the metadata and data values with the NCSA Mosaic program and Vshow utility. The Mosaic and Vshow programs were used to examine the metadata attributes and the structure of the Vgroup. They were also used to examine the name, size, class and field names of the 'Geolocation' and the ‘SwathList’ Vdatas. The Vshow utility was used to examine the actual data values within the data Vdata. The examined data values were compared with the data values in the native TOPEX/Poseidon MGDR read software output.

4.4.7 Migration Time Estimates

Figure 4-4 summarizes the amount of time taken for each of the tasks described above. The total number of person days represented here is the total time taken for the various tasks.

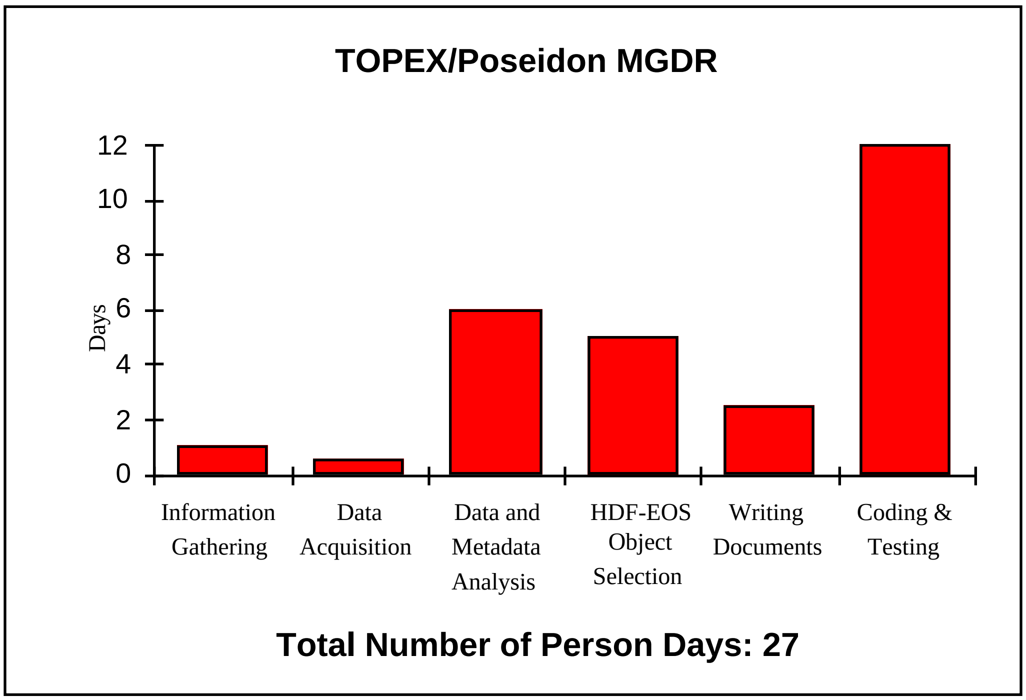

Figure 4-4. Time Taken for the Migration Tasks: TOPEX /Poseidon MGDR Data.

4.5 UARS SOLSTICE

4.5.1 Information Gathering

This task consisted of collecting the information that describes the native format of the UARS SOLSTICE data and metadata. Information gathering can take a long time; however, in this case, the information had already been put on a readily-accessible disk by the ECS DAAC Science Liaison at the GSFC DAAC.

4.5.2 Data Acquisition

This task consisted of collecting a few granules of data to be converted to HDF and collecting any read software for the data files. Again, the data and read software had been made readily-accessible by the ECS DAAC Science Liaison at the GSFC DAAC. The time to acquire the data granules together with the time to gather the format description information was less than one hour. An additional three UARS SOLSTICE granules of data were ordered from the GSFC DAAC. The order was filled in about 1.5 person days.

4.5.3 Data and Metadata Analysis

This task consisted of reading the metadata and data description documents and building and running the data read software.

1) Reading metadata and data description documents**.** This task consisted of reading the format description documents and understanding the metadata and data representations in the data granule. The documents were analyzed (along with the output from the read software described below) in about 2 person days.

2) Building and running of data read software**.** This task consisted of compiling and linking the given read software source code and then running the created software on the sample data granules. The output was compared with the documented data format to help understand how the data are represented in the data granule. Also, an HSTX-developed program was used to display a hexadecimal and ASCII representation of the data to see exactly what the data file contained. The read software was built in less than 1 hour.

4.5.4 HDF-EOS Object Selection

This task consisted of choosing which HDF-EOS objects would be best to represent the various data and metadata in a UARS SOLSTICE granule. The HDF documentation and the draft HDF-EOS Primer were examined to determine which HDF-EOS objects were appropriate. This task took approximately 2 person days.

4.5.5 Writing Documents

This task consisted of writing a preliminary Software Interface Specification (SIS) describing the input data format and the output HDF-EOS file (See Appendix E). Some implementation details such as the names and classes of objects were firmed up during the writing. User Documentation was written to describe the compilation and use of the conversion program (See Appendix J). The writing took approximately 5.5 person days.

4.5.6 Coding and Testing

This task consisted of implementing the design described in the SIS and testing the implementation. The HDF-EOS conversion program was written in ANSI C on a Sun SPARC/20. The testing was done using NCSA's Mosaic and Vshow programs. The HDF file data values displayed by Vshow were compared with the original data values displayed by the data read software to ensure that the values had not changed. The coding and testing took approximately 9 person days. After the HDF file is created, the data within the HDF file must be validated to ensure that no data values have been lost and that no data values have changed or lost precision. Two types of data validation were used: file size examination and data value examination.

4.5.6.1 File Size Examination

File size examination consisted of comparing the size of the HDF file to the sizes of the input UARS SOLSTICE files (see Table 4-5). If the HDF file size is significantly different from the input file sizes, the difference must be explained. In this case, the HDF file is over 5,000 bytes smaller than the input file sizes. This is because the Standard Formatted Data Unit (SFDU) records in the input data file contain a large amount of unused space which is not required in the HDF file.

Table 4-5. UARS SOLSTICE Data File Size vs HDF File Size Comparison.

Data + Metadata File Size	Output HDF File Size	Size Difference	% Difference
13,151 bytes	7,742 bytes	-5,409 bytes	-41.13%

4.5.6.2 Data Value Examination

Data value examination consisted of examining the metadata and data values with both the NCSA Mosaic program and the Vshow utility. The Mosaic program was used to examine the metadata attributes. It was, also, used to examine the names, sizes, classes and field names of the data and parameter Vdatas. The Vshow utility was used to examine the actual data values within the data and parameter Vdatas. The examined data values were compared against the data values in the native UARS SOLSTICE read software output.

4.5.7 Migration Time Estimates

Figure 4-5 summarizes the amount of time taken for each of the tasks described above. The total number of person days represented here is the total time taken for the various tasks.

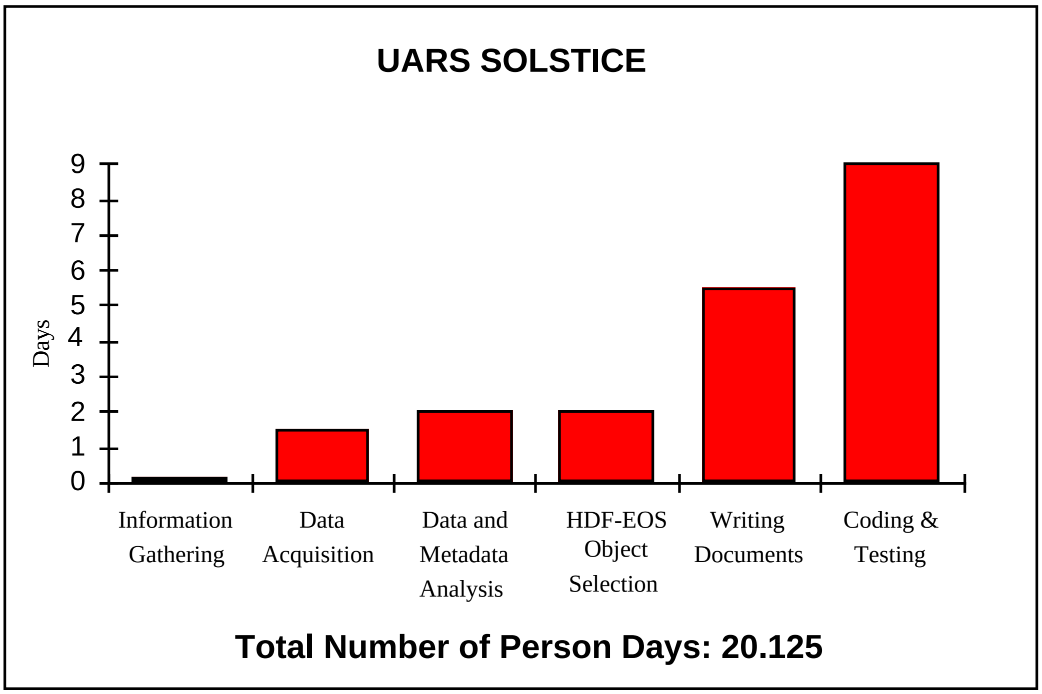

Figure 4-5. Time Taken for the Migration Tasks: UARS SOLSTICE Data.

This page intentionally left blank.

5. Lessons Learned from the Data Migration

Detailed below are some of the product specific lessons we learned from the pilot data migration:

5.1 AVHRR Global 1 Km Land 10 Day Composite

1. The transfer of full-resolution AVHRR Global 1Km Land 10 Day Composite images over the World Wide Web is possible but probably not the best method. One band (Channel 1) of a full-resolution AVHRR 1km image was transferred in approximately 16 hours under near ideal conditions. Requesting a tape wouldn't take much longer and would be more reliable. Also, you could get the entire granule at once instead of 1 band at a time.
2. The AVHRR Global 1Km Land 10 Day Composite metadata is stored in a separate database. This makes it more difficult to acquire all of the pieces (metadata and data) necessary to convert data granules from native format to EOS-HDF. The metadata should be made available in an electronically readable file accessible with the data.
3. The AVHRR Global 1Km Land 10 Day Composite data have to be validated by automated means. This would be true for any large volume of data.
4. When converting large data products to HDF, the input and output files should be on different physical disks. This reduces the disk head movement and should reduce the amount of time needed for conversion.

5.2 FIRE-NWS_In_Sonde

1. Converting FIRE_NMW_IN_SONDE to HDF-EOS didn’t improve the portability of the data since the data were already in the ASCII format.
2. The data values in ASCII were converted to binary in order to provide uniformity across the migration efforts. Conversion of ASCII to binary could be an issue for the larger migration effort.

5.3 GPS Sea Ice Motion Vectors

1. Approximately 10% of the lines of code in the GPS Sea Ice Motion Vectors conversion program were taken without modifications from the AVHRR Global 1 Km Land 10 Day Composite-to-HDF-EOS conversion program. Most of the duplicate lines of code deal with opening and closing disk files and HDF interfaces.

5.4 TOPEX/Poseidon Merged Geophysical Data Record (MGDR)

1. The HDF-EOS specified Swath Data object was the closest in definition to the TOPEX/Poseidon MGDR data.

5.5 UARS SOLSTICE

1. The documentation was not completely accurate. Some fields in the data file did not contain the values that were in the documentation. Also, the data format allows multiple data records, yet, none of the native UARS SOLSTICE granules examined contained multiple data records.
2. When converting an ASCII floating point number to binary, you have to check the resulting value to make sure that you have not lost precision.
3. The record formats of the UARS data products seem similar enough that a single HDF-EOS conversion program could be written which would handle all of them.

6. HDF-Related Problems

6.1 Problems

The following problems were encountered with the 3.3 R4 and 4.0 beta 1 versions of the HDF library and the proposed HDF-EOS standards during the data migration:

1. The data migration will use whatever is the current HDF version at the time of migration. But, the HDF library will continue to evolve to higher versions. So far, the backward compatibility has been maintained; however, evolution of HDF will add more functionality and the earlier versions of HDF will be out of date. This is true with any software package. The backward compatibility of HDF must be ensured during the development of conversion tools and utilities, so that the migrated data will continue to work with new versions of HDF. The relationship between HDF-EOS and HDF is not clear to many users. Similarly, the development of HDF-EOS should be compatible with the newer versions of HDF.

2. An AVHRR Global 1Km Land 10 Day Composite granule is too large (10 GB) to fit within a single HDF file. Ten separate HDF files would be needed to contain a single granule. This is not an HDF problem, strictly. This is a 32 bit machine architecture limitation. Currently, none of the standard data formats, including HDF, support the 64 bit system architecture. But, it is an important requirement that HDF support the handling of large data products. Currently, this problem can be addressed by defining small (less than 2 GB) granules. Having a smaller granule size will also be advantageous for the users of the data.

3 The documentation for HDF is not up to date with the current HDF software. Using some of the newer HDF interfaces was made more difficult because it was hard to find examples of how to use them correctly. An effort should be made to help the HDF documentation keep pace with the HDF software.

4. The HDF 3.3 R4 Vshow source code required changes in order to handle all of the 22 fields in the FIRE-NWS_In_Sonde Pointlist Vdata.

5. The HDF 4.0 B1 library and Vshow source code can handle up to 50 Vdata fields but required changes in order to handle all of the 124 fields in the TOPEX/Poseidon MGDR Swathlist Vdata.

6. The NCSA Collage and Mosaic tools were not always able to handle the HDF files created. There should be greater coordination at NCSA between the development efforts of HDF and the software tools to ensure that HDF and the software tools will work together and that problems in common can be corrected at the same time.

7. The HDF-EOS Point and Grid data object definitions need to be further refined. The required fields and formats need to be specified.

8. The HDF-EOS specified Point data object was the closest in definition to the FIRE-NWS_In_Sonde data. The FIRE-NWS_In_Sonde data matched the Point data object definition but not completely. There were some discrepancies.

9. The UARS SOLSTICE data did not fit any of the three proposed HDF-EOS data objects (Grid, Point and Swath). Currently, we think it fits the HDF Science Data Table object. Therefore, the Grid, Point and Swath data objects may not be appropriate for all data products.

10. There need to be some guidelines for determining which HDF-EOS data object is most appropriate for a given data product. For example, the GPS Sea Ice Motion Vectors data product is derived from 2 gridded images. It is almost, but not quite, a gridded product (not every grid element is represented). It could also be put into a Point data object but it did not seem quite appropriate because there are two sets of latitudes and longitudes.

6.2 Recommendation

We recommend that NCSA and HDF-EOS development efforts be made one of the high priorities for ESDIS and ECS before the Engineering Phase of the Data Migration begins. This will result in cost and resource savings for the Engineering Phase.

7. Migration Time Estimates

7.1 Estimated Time by Migration Task

We have identified the following 6 tasks for the Pilot Data Migration: information gathering, data acquisition, data and metadata analysis, HDF-EOS object selection, writing documents, and coding and testing. Figure 7-1 shows the time (in person days) taken for the various conversion tasks.

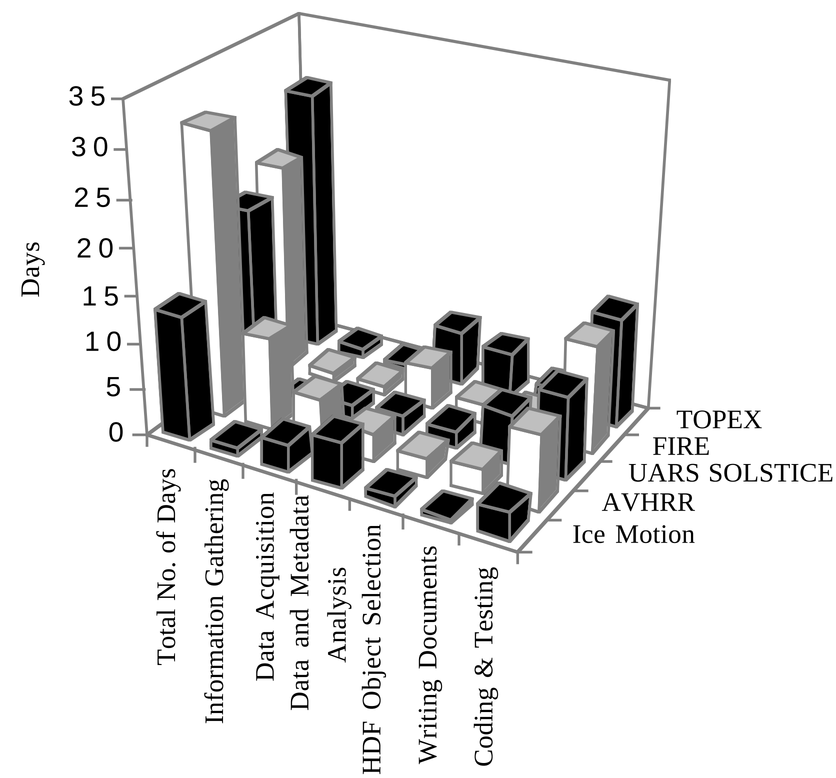

Figure 7-1. Time for Migration by Data Product and by Conversion Task

After examining the figure, there seem to be certain trends. Coding and testing always takes a large amount of the time and is usually, but not always, the largest part of the time. However, other tasks should not be minimized. The information gathering step has the widest time variance. This is a direct result of the media of the information, either electronic or printed. The printed material (documentation) always seems to take more time to acquire. The data acquisition time averages out to about 2 person days with much less variation than information gathering. This is because the data are electronically readable and therefore easier to access. Information gathering is a potential bottleneck for Version 0 migration because of the large number of Version 0 data products and their associated information. Extensive cooperation by the DAACs will certainly help mitigate this potential problem.

The time taken for the HDF-EOS object selection was just about the same for every product, except TOPEX/Poseidon MGDR, which had a complex data structure. This implies that the time needed to select HDF-EOS objects for various data products won't vary much after an HDF object selection criteria has been developed. The time to write various documents was also about the same for every data product, except for UARS SOLSTICE. This is because the UARS SOLSTICE data product was the first data product converted. The format and contents of the documents had to be defined before the writing could take place for the UARS SOLSTICE data product.

Over all, the combination of the data and metadata analysis task, and coding and testing task, did take more time than all of the other tasks combined. The other tasks each took approximately equivalent amounts of time with writing documents taking slightly longer.

Four people worked on the study but the actual coding was done by two programmers. One of them was new to HDF and the other programmer had about 8 months experience with HDF. It is not possible to determine if the programmer profile had any impact on the time taken to migrate a data product. The sample size of two programmers is too small to produce any statistically significant inference.

Additional factors that may influence operational time are the processes involving formal review and other configuration control procedures. The code was generated, tested and validated with a low level effort. The time required to migrate a data product is largely dependent on the complexity of the data, availability of the documentation, metadata and other associated information.

The data and metadata analysis task averages out to about 4 person days and is frequently the most lengthy task except for coding and testing. This is not a profound observation. But, it does support the conclusion that analysis, coding and testing are the major tasks of any software development, including data migration.

7.2 Estimated Time by Data Product

Figure 7-2 summarizes the time (in person days) needed to convert the various data products from their native formats to HDF-EOS.

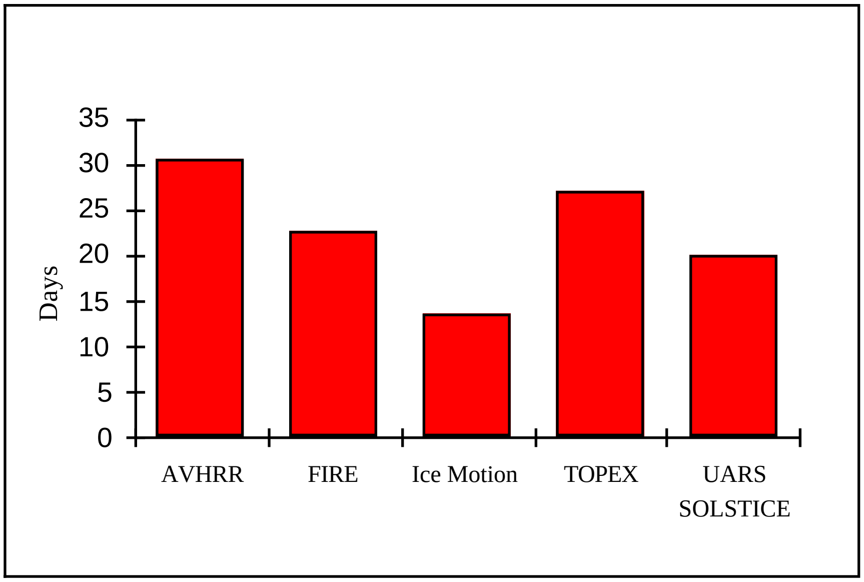

Figure 7-2. Time Taken for the Migration of each Data Product

The data in the above figure should not be taken to mean that the migration of a data product similar to one of the data products converted here will take an equivalent amount of time. The availability of information and data, which can have a large impact on the conversion time, varies widely. In some cases, information was available on-line and hence took less time. The organization of data and metadata was also a factor. The availability of good specification documents can reduce the analysis time.

No formal quality control, testing or validation procedures were used during this pilot migration. These procedures will add time to the conversion of any data product. In some cases, there was a time overlap between the various conversion tasks for a product. This necessitated estimating the amount of time taken for each task which may not be the same as the actual amount of time taken. Also, a particular data product was not necessarily converted over a sequential number of person days. There were times when other tasks intervened.

Based on our experience from this study, it will take 2 to 3 months to develop and test prototype software, or 3 to 4 months to develop and test operational software, to convert one V0 data product to HDF-EOS.

This page intentionally left blank.

Appendix A
AVHRR Global 1 KM Land 10 Day Composite to HDF-EOS Software Interface Specification

Description of the Native Advanced Very High Resolution Radiometer (AVHRR) Global 1 Km Land 10 Day Composite Data Product

The AVHRR Global 1-Km Land 10 Day Composite data product is composed of data acquired by the National Oceanic and Atmospheric Administration's (NOAA) Polar Orbiting Environmental Satellites (POES). The Data Product Reference number for the AVHRR Global 1-Km Land 10 Day Composite data product is E-2 in the Science Data Plan (SDP) dated July 1994. The data can be obtained from the EROS Data Center (EDC) Distributed Active Archive Center (DAAC) at the following World Wide Web address:

http://sun1.cr.usgs.gov/landdaac/1KM/comp10d.html.

The data are composed of 5-channel, 10-bit, raw AVHRR images, at 1.1-km resolution (at nadir) for every daily afternoon pass over all land and coastal zones using data from NOAA's polar-orbiting NOAA-11 satellite. Each image consists of 17,347 lines with each line containing 40,031 pixels. The data collection began April 1, 1992.

The data product is in a native format and contains the following 10 bands:

Band 1 is the AVHRR channel 1 data (visible spectrum). Band 2 is the AVHRR channel 2 data (near-infrared spectrum). Bands 3 through 5 are the AVHRR channel 3 through 5 data (all thermal spectrum channels). Band 6 is the normalized difference vegetation index (NDVI) data. Band 7 is the satellite zenith data. Band 8 is the solar zenith data. Band 9 is the relative azimuth data. Band 10 is the date index data.

The first 5 bands each have a size of approximately 1324 Megabytes (MB). The last 5 bands each have a size of approximately 662 Megabytes (MB). The total size for a granule is approximately 10 Gigabytes (GB).

Proposed Design Plan

Due to the large volume of data, we converted only band 6 (NDVI) to HDF. The metadata attributes fit best in the HDF-EOS datatype 'P=V Metadata'. The actual NDVI data values fit best in the HDF-EOS datatype 'Grid'. The HDF-EOS datatypes are specified in the Draft HDF-EOS Primer for Version 1 EOSDIS (175-WK-001-001) dated January 1995.

The metadata will be stored in the HDF file as attributes. The Grid data object will consist of a Vgroup containing a geometry Vdata and a data SDS. The Vdata will contain projection information. The SDS will contain the actual data values. Figure A-1 depicts the proposed design plan.

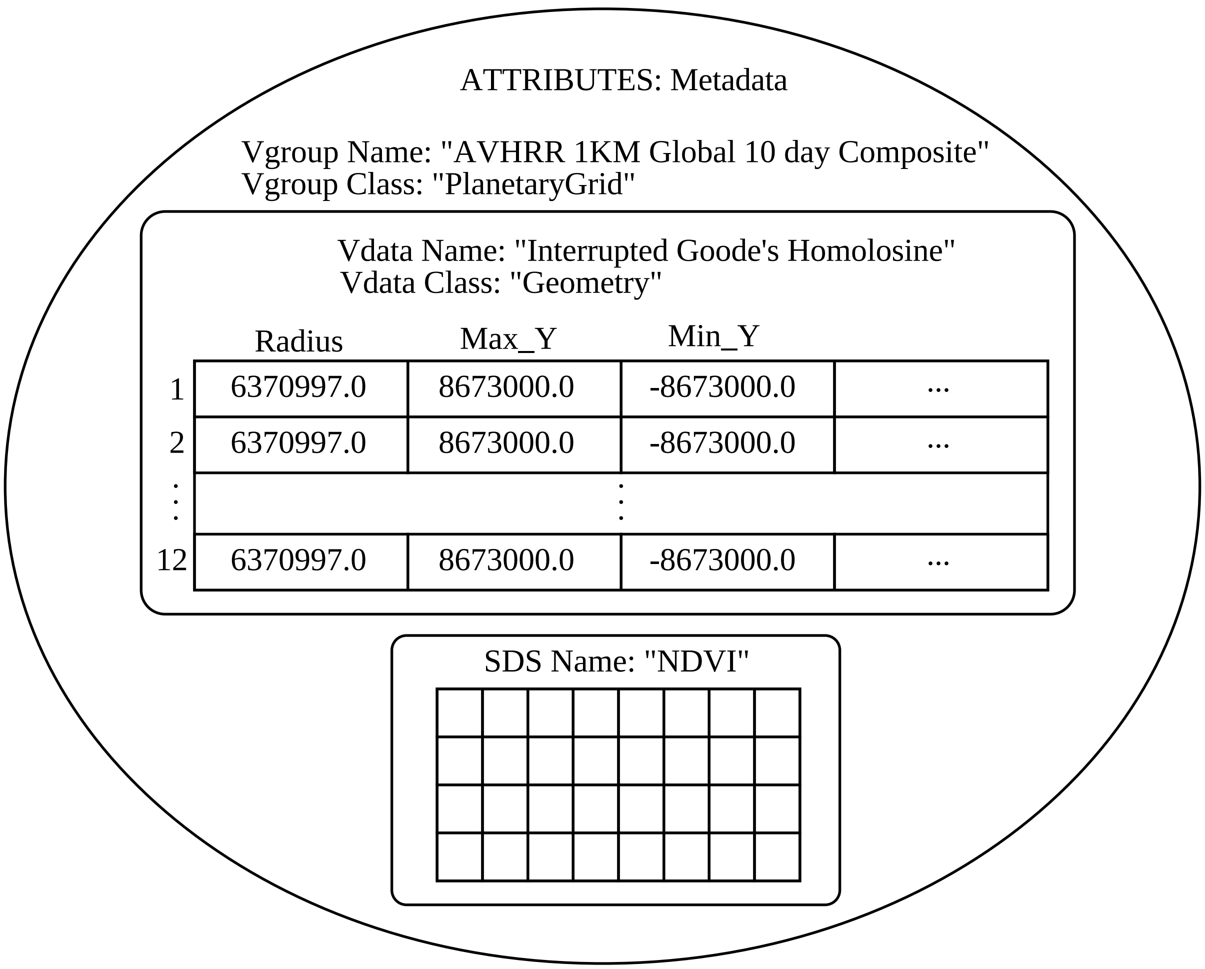

Figure A-1. Design Plan for the HDF-EOS Implementation of the AVHRR Global 1 Km Land 10 Day Composite Data Product.

Description of HDF file

The metadata for the AVHRR granules is stored in a separate database at EDC. The metadata for this particular granule has been extracted from the database and manually placed in an ASCII file for use by the HDF conversion program. All of the metadata will be stored as global attributes within the HDF file.

The NDVI data will be read from a binary file and stored in an SDS named 'NDVI' within the HDF file.

The map projection parameters will be stored in a Vdata named "Interrupted Goode's Homolosine" with a class of "Geometry" within the HDF file. There will be an entry within the Vdata for each of the 12 geographic regions of the data. The following is a list of the parameters stored within the Vdata:

1.	Radius of Sphere	(meters)
2.	Maximum Y	(Goode's meters, top of image)
3.	Minimum Y	(Goode's meters, bottom of image)
4.	Minimum X	(Goode's meters, left side of image)
5.	Maximum X	(Goode's meters, right side of image)
6.	Resolution	(meters)
7.	Registration	(Center)

The following parameters differ for each region:

8.	Geographic Region	(range 1 to 12)
9.	Northern-most Latitude	(radians, top of region)
10.	Southern-most Latitude	(radians, bottom of region)
11.	Western-most Longitude	(radians, left side of region)
12.	Eastern-most Longitude	(radians, right side of region)
13.	Projection	(Mollweide or Sinusoidal)
14.	Central Meridian	(radians)

15.	False Easting	(radians)
16.	Goode's Window Start Line	(range 1 to 17347)
17.	Goode's Window Start Sample	(range 1 to 40031)
18.	Goode's Window Number of Lines	(range 1 to 17347)
19.	Goode's Window Number of Samples	(range 1 to 40031)
20.	Local Projection Y	(meters, top of region)
21.	Local Projection X	(meters, left side of region)
22.	Goode's Projection Y	(Goode's meters, top of region)
23.	Goode's Projection X	(Goode's meters, left side of region)

A Vgroup named "AVHRR 1KM Global 10 Day Composite" with a class of "PlanetaryGrid" will contain the Geometry Vdata and the Data SDS within the HDF file.

Implementation

The HDF library version 3.3 release 4 will be used for the conversion of the AVHRR data to HDF. The computer platform used will be a Sun SPARC/20 workstation.

Additional Information

The metadata associated with the data consists of 12 fields which includes the beginning date of the composite, the ending date of the composite, a geographic region code, a 10 day period number, the source satellite number, a band description and a few additional fields.

Appendix B
FIRE-NWS_In_Sonde to HDF-EOS
Software Interface Specification

Description of the Native FIRE-NWS_In_Sonde Data Product

The NWS_IN_SONDE (National Weather Service Inner-network Rawinsonde) data are a First ISCCP Regional Experiment (FIRE) Cirrus II data product. The FIRE-NWS_In_Sonde data were collected using the Rawinsonde sensor from ground stations. The Data Product Reference number for FIRE Cirrus II, NWS_IN_SONDE data is L-55 in the Science Data Plan (SDP) dated July 1994.

The FIRE-NWS_In_Sonde data product has been formatted in ASCII. These data have been collected from various stations for the period between Nov. 13, 1991 to Dec. 7, 1991. The FIRE-NWS_In_Sonde data consists of 17 granules and each granule has multiple ASCII data files. The first ten lines in a data file contain header information, followed by a two-line variable name heading, followed by the NWS sondes records. Each record starts with the number of the level, followed by 21 variables, and ends with the same number of the level. These variables are separated by white spaces and will be described in detail later.

The FIRE-NWS_In_Sonde data could vary in size according to the number of ASCII files present in each granule. The size of a single ASCII file is around 50K bytes. The translation program will work with all of the ASCII files within a single granule. Each execution of the translation software will generate a single HDF file for a single ASCII input file. This implies that there would be the same number of HDF files generated as there are input files.

The above information has been extracted from a readme file provided by the LaRC DAAC User and Data services (UDS) office. For more information contact:

Langley DAAC User and Data Services Office
Phone Number: (804) 864-8656
E-mail address: userserv@eosdis.larc.nasa.gov

Proposed Design Plan

The FIRE-NWS_In_Sonde data have been formatted in ASCII. These data are similar to the Point Data structure as defined in the HDF-EOS Primer (Draft Version) for Version 1 EOSDIS, January 1995.

The FIRE-NWS_In_Sonde data are similar to the Moving Platform Point Data structure. The HDF file will contain the mandatory Index Vdata and a Data Vdata which will contain all the data parameters. The Vgroup will be named FIRE-NWS_In_Sonde with appropriate dates and will have the class Pointdata. The header record of the input FIRE-NWS_In_Sonde file will be stored as attributes. Figure B-1 depicts the proposed design plan for the FIRE-NWS_In_Sonde data.

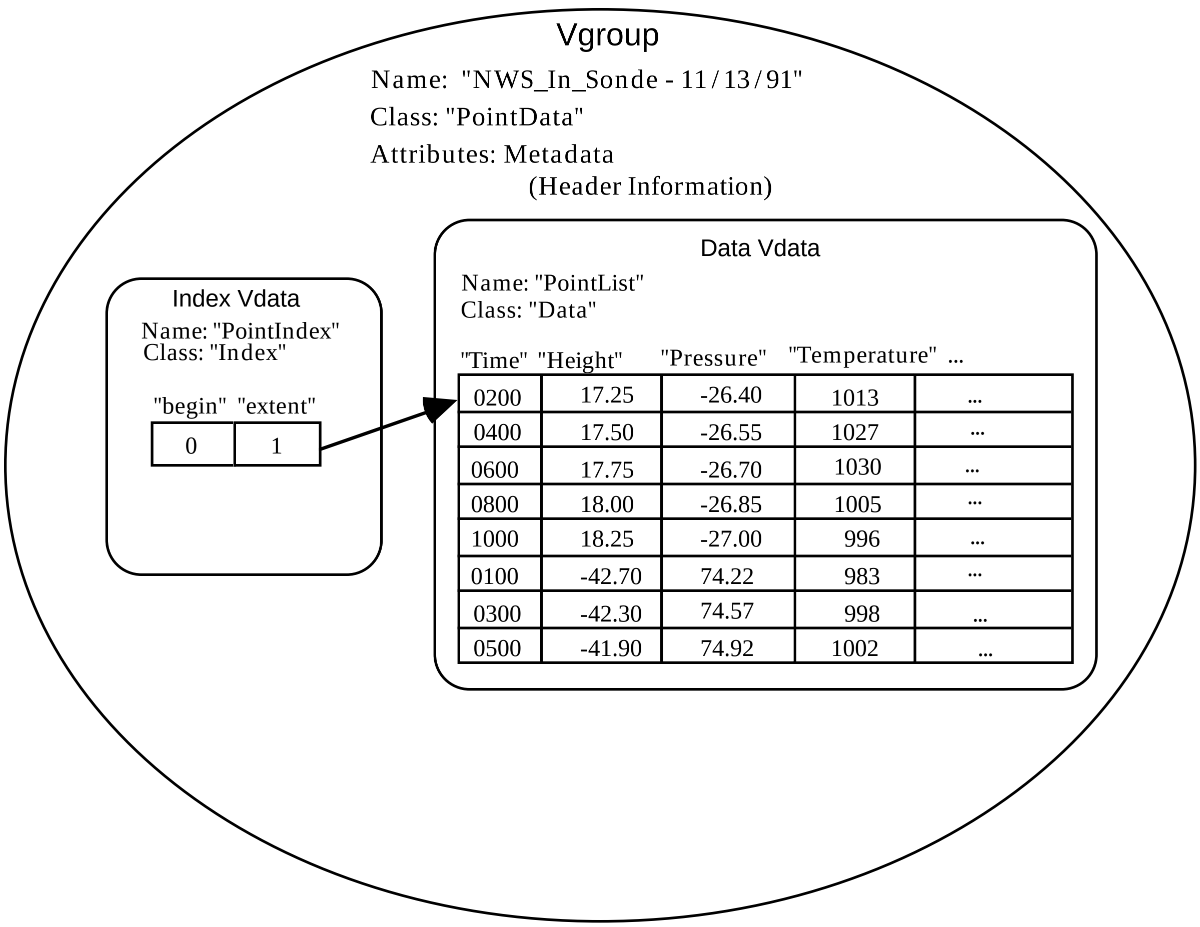

Figure B-1. Design Plan for the HDF-EOS Implementation of the FIRE _NWS_IN_SONDE Data Product.

Description of the HDF file

The HDF file will contain one Vgroup. This Vgroup will be named NWS_IN_SONDE with appropriate dates and will have the class Pointdata. The header record will be stored as attributes. The Vgroup will contain two Vdatas namely, Index Vdata and Data Vdata. The Index Vdata will have the mandatory begin and extent fields. The Data Vdata will contain all the data variables from the input file.

The following is a list of the data variables with their units, in the order of their occurrence:

Variable Name

1. time into launch (minutes)
2. height (km)
3. pressure (mb)
4. temperature (C)
5. theta - potential temperature (K)
6. rh - relative humidity with respect to water (percent)
7. rh - relative humidity with respect to ice (percent)
8. sphum - specific humidity (g/kg)
9. dewpt - dew point temperature (C)
10. frstpt - frost point temperature (C)
11. speed - wind speed (m/s)
12. direc - wind direction (degrees)
13. u - zonal component of wind (m/s)
14. v - meridians component of wind (m/s)
15. ascent - ascent rate (C/km)
16. lapse - lapse rate (C/km)
17. dtheta - potential temperature lapse rate (K/km)
18. x - longitudinal distance of sonde with respect to point of origin (km)
19. y - latitudinal distance of sonde with respect to point of origin (km)
20. lat - latitudinal position of sonde (degrees N)

21. long - longitudinal position of sonde (degrees W)

Implementation

The HDF library version 3.3 release 4 will be used for the conversion of the FIRE-NWS_In_Sonde data to HDF. The computer platform used will be a Sun SPARC 20 workstation.

Appendix C
GPS Sea Ice Motion Vectors to HDF-EOS Software Interface Specification

Description of the Native GPS Sea Ice Motion Vectors Data Product

The GPS Sea Ice Motion Vectors data product is derived from data acquired by the Synthetic Aperture Radar (SAR) instrument on the European Space Agency's (ESA) first European Remote Sensing satellite (ERS-1). The Data Product Reference number for GPS Sea Ice Motion Vectors is A-12 in the Science Data Plan (SDP) dated July 1994. The data are available from the Alaska SAR Facility (ASF) Distributed Active Archive Center (DAAC).

This data product was generated from pairs of geocoded low resolution SAR imagery. The data represent how far ice features move and through what angle they rotate between images. The time difference between repeat viewing of the same ice field depends on the orbit characteristics of the ERS-1 platform and the geographic location of the region of interest. The typical values are between 3 and 35 days. The absolute spatial displacement of ice features is measured and reported for every cell of the 5 km SSM/I grid covered by the selected pair of images which had detectable ice motion. Each ice motion vector product covers approximately a 100 km x 100 km area. This level 3 product contains an ice feature's initial latitude, initial longitude, final latitude, final longitude, X displacement in kilometers, Y displacement in kilometers, rotation angle, and a reliability measure. The data are presented in the CEOS format.

The above information has been extracted from the Global Change Master Directory (GCMD) Directory Interchange Format (DIF) and the Guide document for the GPS Sea Ice Motion Vector data product. The GPS Sea Ice Motion Vector data products consist of three separate files: a leader file, a data file, and a trailer file. Each file contains a mixture of ASCII and binary values. The leader and trailer files are each approximately 1,300 bytes long. The data file is approximately 25,000 bytes long.

For additional information contact:

ASF User Services / Alaska SAR Facility

Phone number: 907-474-6166

E-mail addresses: asf@eos.nasa.gov or uso@eosims.asf.alaska.edu

Proposed Design Plan

The GPS Sea Ice Motion Vectors native format consists of CEOS formatted records. The metadata attributes fit best in the HDF-EOS datatype 'P=V Metadata'. The actual GPS Sea Ice Motion Vectors data values fit best in the HDF-EOS datatype 'Science Data Table'. The HDF-EOS datatypes are specified in the Draft HDF-EOS Primer for Version 1 EOSDIS (175-WK-001-001) dated January 1995.

The GPS Sea Ice Motion Vectors metadata will be stored in the HDF file as attributes. A Vdata (Science Data Table) object will contain the actual data values. Figure C-1 depicts the proposed design for the migration of the GPS Sea Ice Motion Vectors data product.

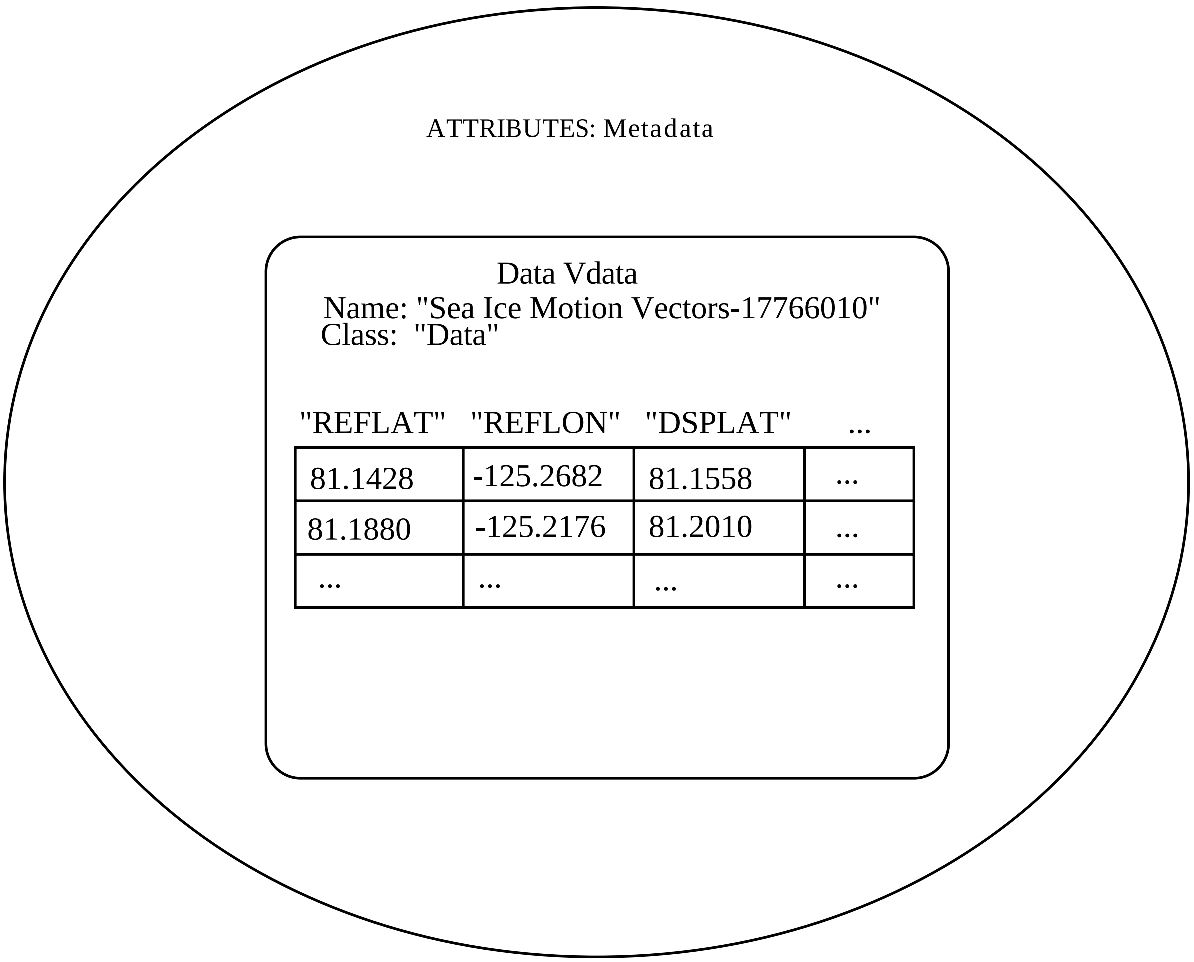

Figure C-1. Design Plan for the HDF-EOS Implementation of the GPS Sea Ice Motion Vectors Data Product.

Description of HDF file

The metadata for each GPS Sea Ice Motion Vectors granule is located in the native format in both the leader and trailer files. All of the metadata will be stored as global attributes within the HDF file.

The data for each GPS Sea Ice Motion Vectors granule consists of ice motion data and quality information associated with the ice motion data. The GPS Sea Ice Motion vectors data will be stored in a Vdata named "Sea Ice Motion Vectors Data - XXXX" where "XXXX" represents the product identifier. The data Vdata will have a column for each of the sea ice motion data parameters and a column for the quality information. The data and quality values will be converted from ASCII to binary numbers before being stored in the data Vdata.

The following is a list of the data and quality fields with their units:

Field Name

1. REFLAT - latitude of grid point, reference image (degrees)
2. REFLON - longitude of grid point, reference image (degrees)
3. DSPLAT - latitude of grid point, displaced image (degrees)
4. DSPLON - longitude of grid point, displaced image (degrees)
5. DELTAX - displacement in X direction (kilometers)
6. DELTAY - displacement in Y direction (kilometers)
7. ROTANGLE - rotation angle (degrees)
8. MATCH - goodness of match metric

Implementation

The HDF library version 4.0 beta 1 release will be used for the conversion of the GPS Sea Ice Motion Vectors data to HDF. The computer platform used will be a Sun SPARC 20 workstation.

This page intentionally left blank.

Appendix D
TOPEX/Poseidon Merged Geophysical Data Record to HDF-EOS Software Interface Specification

---

Description of the Native TOPEX/Poseidon Merged Geophysical Data Record (MGDR) Data Product

Merged GDR data products are generated from TOPEX/Poseidon measurements. These data products have been designed to reinforce unity of the TOPEX/Poseidon mission. TOPEX data are available over both land and oceans, whereas Poseidon altimeter data are available only over oceans. The data products consist of 10-day repeat cycles of this data and are designed to be as homogenous as possible so that they can be used for ocean and geophysical studies. Basically, MGDR data includes the measurement locations based on orbit ephemeris, altimeter height measurements and associated corrections. The Data Product Reference number for the MGDR data is J-32 in the Science Data Plan (SDP) dated July 1994.

The MGDR data are grouped and organized with a single cycle header file and a maximum of 254 pass files for every cycle. The cycle header file is formatted in ASCII and provides organization and data product identification for the 10-day repeat cycle. A single pass file is considered to be a granule of the data product. A pass file contains altimeter data from a satellite pass (half a revolution). Each pass file has a header part and a data part. The header part is formatted in ASCII and provides information about the product, calibration results, orbit quality and typical pass characteristics. The data part is formatted as a VAX binary integer type. The data part is a time record and has N scientific data records (N <= 3360). A scientific data record contains 124 fields, each stored as one, two or four bytes. The size of each MGDR pass file is approximately 1,100,000 bytes. The translation program will work with a single pass file (granule) at a time. Each execution of the translation software will generate a single HDF file for a single MGDR pass file. This implies that there would be the same number of HDF files generated as there are input pass files.

The above information has been extracted from the Merged GDR (TOPEX/Poseidon) Users Handbook dated September 13, 1993. This handbook was part of the information provided in a Compact Disc by the Jet Propulsion Laboratory. This CD contains 2 cycles of MGDR data, the handbook, etc. For more information contact:

Jet Propulsion Laboratory
Phone Number: (818) 354-9890

E-mail address: podaac@podaac.jpl.nasa.gov

Proposed Design Plan

The MGDR data have metadata formatted in ASCII and the data formatted in a VAX binary integer type. These data are similar to the Swath Data structure as defined in the HDF-EOS Primer (Draft Version) for Version 1 EOSDIS, January 1995.

The HDF file will contain two Vdatas, one Vgroup and the header record will be stored as Global Attributes. The first Vdata will contain the data parameters and will be called Data Vdata. The second Vdata will be a table containing the Geolocation information and will be called the Geolocation Vdata. The Vgroup will be named MGDR (TOPEX/Poseidon) and will have the class SwathData. Figure D-1 depicts the proposed design for the migration of the TOPEX/Poseidon MGDR data product.

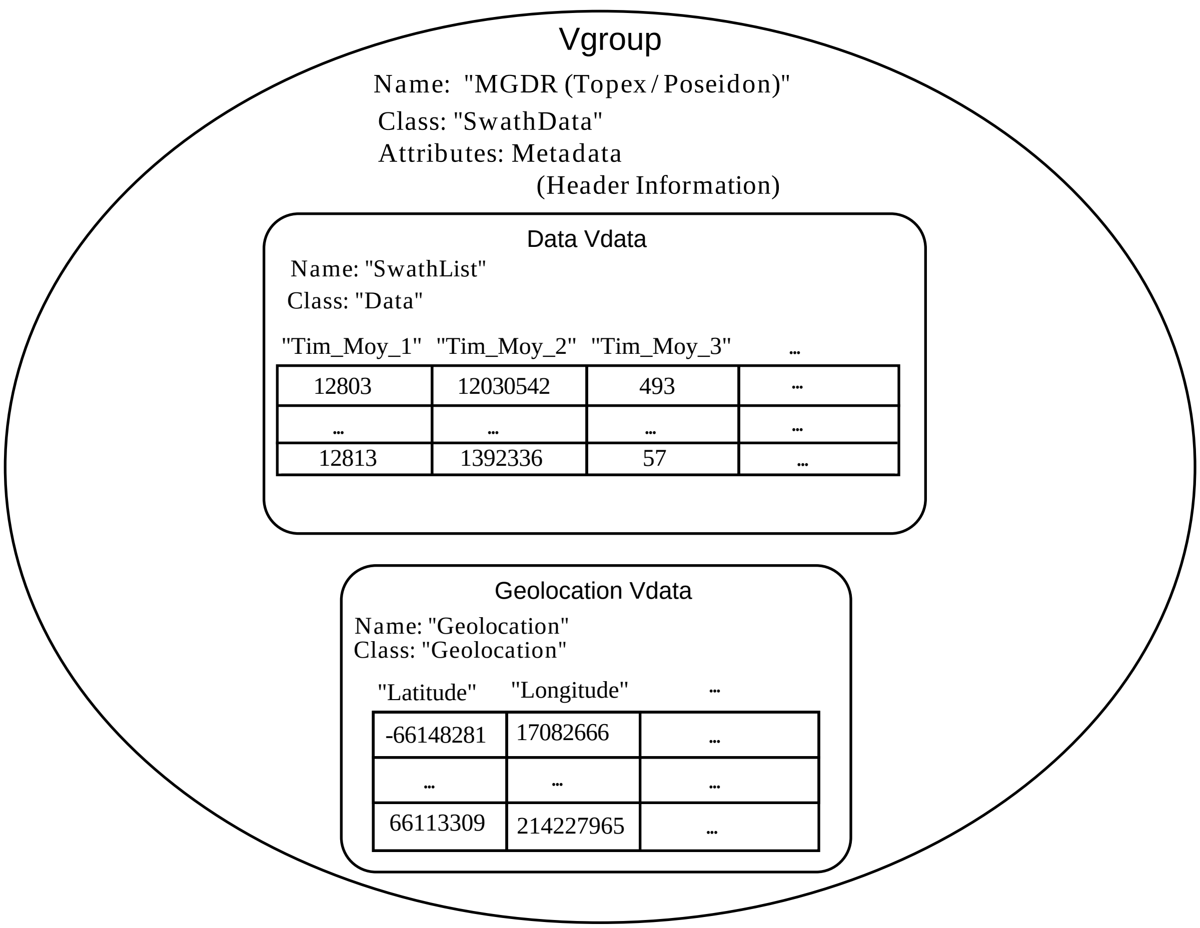

Figure D-1. Design Plan for the HDF-EOS Implementation of the TOPEX /Poseidon MGDR Data Product.

Description of the HDF file

The HDF file will contain one Vgroup. This Vgroup will be named MGDR (TOPEX/Poseidon) and will have the class SwathData. The header record will be stored as global attributes. The Vgroup will contain two Vdatas namely, Data Vdata and Geolocation Vdata. The Data Vdata will contain all the data parameters from the input file. There are around 124 data parameters in the input file. The Geolocation Vdata will be a table containing the geolocation information.

Implementation

The HDF library version 3.3 release 4 will be used for the conversion of the TOPEX/Poseidon MGDR data to HDF. The computer platform used will be a Sun SPARC 20 workstation.

This page intentionally left blank.

Appendix E
UARS SOLSTICE to HDF-EOS
Software Interface Specification

Description of the Native UARS SOLSTICE Data Product

The Solar-Stellar Irradiance Comparison Experiment (SOLSTICE) is one of the instruments collecting data on the Upper Atmosphere Research Satellite (UARS). The Data Product Reference number for UARS SOLSTICE is G-15 in the Science Data Plan (SDP) dated July 1994.

The UARS SOLSTICE instrument measures the energy input to the upper atmosphere by observing the solar spectral irradiance in the ultraviolet spectral range from 119 to 420 nm. The UARS SOLSTICE measures the full disk solar irradiance with high precision and accuracy to follow short-term (minutes to hours), intermediate-term (days to weeks), and long-term (11 year sunspot and 22 year solar magnetic field cycles) variations in the solar output. The UARS SOLSTICE instrument is a three-channel ultraviolet spectrometer which allows for observations of both the sun and bright blue stars using the same optics.

The UARS SOLSTICE instrument is mounted on the two-axis Solar Stellar Pointing Platform (SSPP) which can track objects to an accuracy of 1.5 arc minutes. During the daylight portion of an orbit UARS SOLSTICE is pointed at the sun, and during the nighttime portion it is pointed toward blue stars. One or more full spectral scans of the sun at the instrument resolution is completed each orbit. Stellar scans require longer integration times, and are only made at a few selected wavelengths throughout the spectral range, with typically one or two stars at one or two wavelengths being observed each orbit.

Two UARS SOLSTICE Level 3 data products are produced each day: Level 3AS and Level 3BS. This document only covers Level 3BS data products.

The Level 3BS daily data product contains a single array representing a daily mean solar spectrum. Each element in the array is spaced at 1.0 nm intervals and centered on the half nm from 115 to 420 nm. The array elements have been normalized to an earth-sun distance of 1 AU. Also produced are daily average integrated intensities (Gaussian fits) or core/wing ratios (Mg-II at 280 nm and Ca-II at 393 nm) for selected solar spectral features. A table (calibration file) of instrument degradation function coefficients which have been determined from selected bright blue stars has been applied to the data. The time-degradation function (which varies with wavelength) has also been applied to the irradiance spectrum and the integrated irradiance of the selected spectral emission features, but not to the core/wing ratios.

The above information has been extracted from the Global Change Master Directory (GCMD) Directory Interchange Format (DIF) for UARS SOLSTICE.

The UARS SOLSTICE Level 3BS data products consist of two separate files: a metadata file and a data file. The metadata file is approximately 700 bytes long. The data file is approximately 12500 bytes long.

For additional information contact:

GSFC/DAAC User Support Office
Phone number: (301) 286-5033
E-mail address: daacuso@eosdata.gsfc.nasa.gov

Proposed Design Plan

The UARS SOLSTICE native format consists of Standard Formatted Data Units (SFDU). The metadata attributes fit best in the HDF-EOS datatype 'P=V Metadata'. The actual UARS SOLSTICE data values fit best in the HDF-EOS datatype 'Science Data Table'. The HDF-EOS datatypes are specified in the Draft HDF-EOS Primer for Version 1 EOSDIS (175-WK-001-001) dated January 1995.

The UARS SOLSTICE metadata will be stored in the HDF file as attributes. A Data Vdata (Science Data Table) object will contain the actual data values. A Parameter Vdata will contain the ASCII data parameter values. Figure E-1 depicts the proposed design for the migration of the UARS SOLSTICE data product.

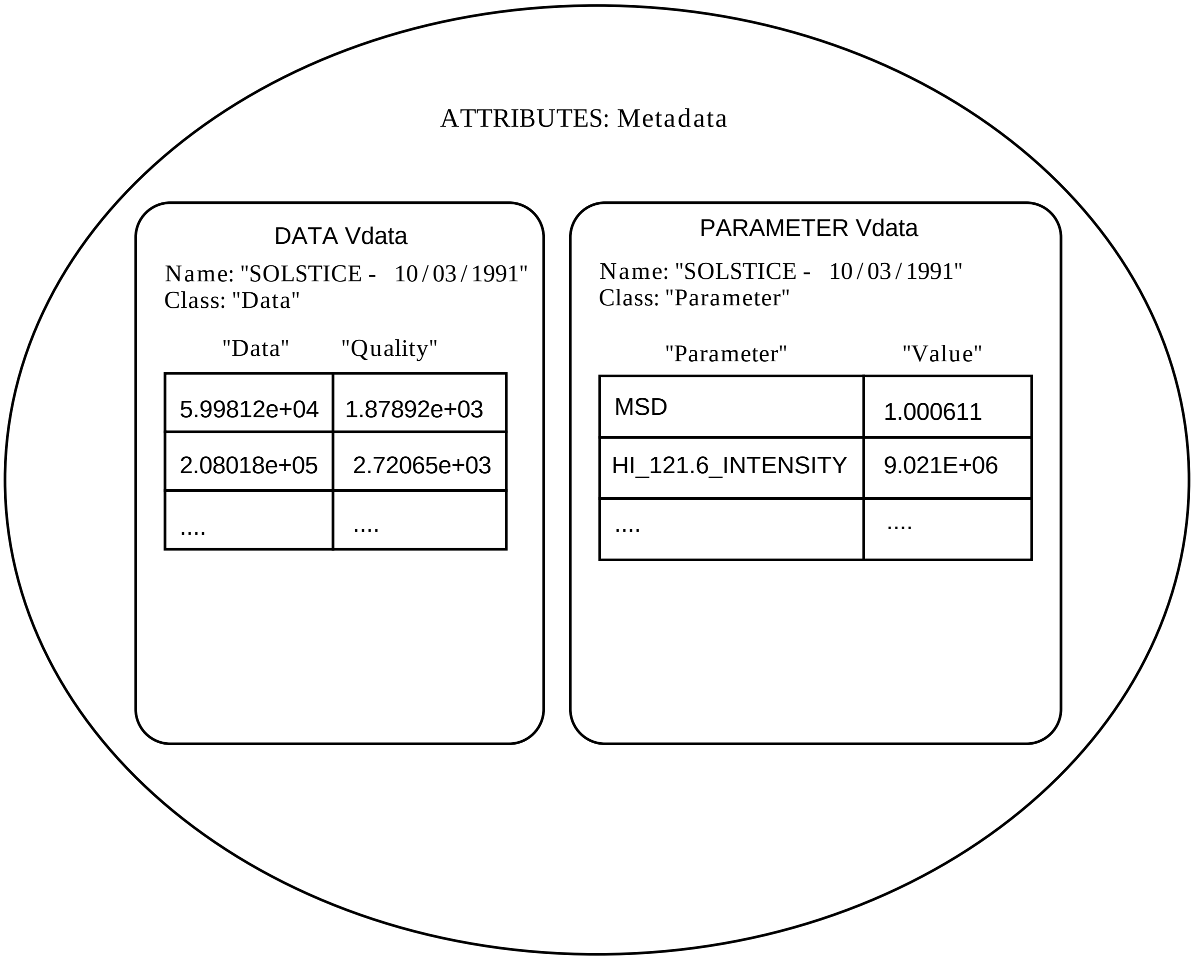

Figure E-1. Design Plan for the HDF-EOS Implementation of the UARS SOLSTICE Data Product.

Description of HDF file

The metadata for each UARS SOLSTICE granule is located in the native format in the following locations: 1) within the UARS SOLSTICE metadata file and 2) within the UARS SOLSTICE data file preceding the actual data. All of the metadata will be stored as global attributes within the HDF file.

The data for each UARS SOLSTICE granule consists of irradiance data and quality information associated with the irradiance data. The UARS SOLSTICE data will be stored in a Vdata named "SOLSTICE Data - MM/DD/YYYY" where "MM/DD/YYYY" represents the date that the data were collected. The data Vdata will have a column for the irradiance data and a column for the quality information. It is indexed by wavelength subtracted from the starting wavelength.

The ASCII data parameters (stored within the data file following the irradiance data values) will be stored in a Vdata named "SOLSTICE Data - MM/DD/YYYY" where "MM/DD/YYYY" represents the date that the data were collected. The parameter Vdata will have a column for the data parameter name and a column for the data parameter value. The following is a list of the data fields with their units:

	Field Name
1.	Data - solar flux measurements in (watt/m^3)
2.	Quality - standard errors associated with data values in (watt/m^3)

The following is a list of data parameters with their units:

	Parameter Name	
1.	Mean Solar Distance	- (AU)
2.	Hydrogen I Emission Intensity	- (watt/m^3)(nm)
3.	Hydrogen I Emission Standard Error	- (watt/m^3)(nm)
4.	Oxygen I Emission Intensity	- (watt/m^3)(nm)
5.	Oxygen I Emission Standard Error	- (watt/m^3)(nm)
6.	Carbon IV Emission Intensity	- (watt/m^3)(nm)
7.	Carbon IV Emission Standard Error	- (watt/m^3)(nm)
8.	Carbon I at 156.1 nm Emission Intensity	- (watt/m^3)(nm)
9.	Carbon I at 156.1 nm Emission Standard Error	- (watt/m^3)(nm)
10.	Carbon I at 165.6 nm Emission Intensity	- (watt/m^3)(nm)
11.	Carbon I at 165.6 nm Emission Standard Error	- (watt/m^3)(nm)
12.	Magnesium II Core to Wing Ratio	- (nm)
13.	Magnesium II Core to Wing Error	- (nm)
14.	Calcium II Core to Wing Ratio	- (nm)
15.	Calcium II Core to Wing Error	- (nm)
16.	Carrington Longitude	- (degrees)
17.	Carrington Latitude	- (degrees)

Implementation

The HDF library version 3.3 release 4 will be used for the conversion of the UARS SOLSTICE data to HDF. The computer platform used will be a Sun SPARC 20 workstation.

This page intentionally left blank.

Appendix F
AVHRR Global 1 Km Land 10 Day Composite
HDF-EOS User Documentation

The 'a2h' program reads AVHRR Global 1 Km Land 10 Day Composite NDVI data files distributed from the EDC DAAC and converts them to HDF formatted files. The inputs consist of a metadata file and a data file.

The program is written in ANSI C and requires the HDF library (3.3r4 or later) in order to compile. If you are using a Sun SPARC workstation with the HDF include libraries in the directory /usr/local/include and the library files (libdf.a and libnetcdf.a) in the directory /usr/local/lib, the following command will create an executable program a2h from the a2h.c source code file:

```
cc -DSUN -I/usr/local/include a2h.c -o a2h -L /usr/local/lib -lnetcdf -ldf
```

When using a machine other than a Sun, replace the -DSUN with the appropriate machine type in the above command line. For example, if you are using a Silicon Graphics machine, replace the -DSUN with -DIRIS4. A list of the supported machine types appears in the file HDF3.3rX/README where 'X' is the current release number.

To run the a2h program, type the program name followed by the metadata file name and the NDVI data file name.

Example: a2h avhrr.meta avhrr-ndvi-060192

The a2h program will create the HDF file in the current directory with a name derived from the data file name. For the above example the HDF file name would be

‘avhrr-ndvi-060192.hdf’.

Note: The example metadata file was created with an editor. The metadata attributes are in the form of parameter = value within the metadata file.

This page intentionally left blank.

Appendix G
FIRE-NWS_In_Sonde
HDF-EOS User Documentation

A single FIRE-NWS_In_Sonde granule consists of approximately 80-100 individual files distributed by the LaRC DAAC. Each execution of the 'nws2hdf' program reads a single file of the granule and then converts it to an HDF formatted file. The input consists of a single file, which contains both the metadata and the data values.

The program has been written in ANSI C and requires the HDF library (3.3r4 or later) in order to compile. If you are using a Sun SPARC workstation with the HDF include libraries in the directory /usr/local/include and the library files (libdf.a and libnetcdf.a) in the directory /usr/local/lib, the following command will create an executable program nws2hdf from the nws2hdf.c and the nwslib.c source code files:

```
cc -DSUN  -I/usr/local/include nws2hdf.c nwslib.c -o nws2hdf -L /usr/local/lib -lnetcdf -ldf
```

When using a machine other than a Sun, replace the -DSUN with the appropriate machine type in the above command line. For example, if you are using a Silicon Graphics machine, replace the -DSUN with -DIRIS4. A list of the supported machine types appears in the file HDF3.3rX/README where 'X' is the current release number.

To run the nws2hdf program, type the program name followed by the data file name.

Example: nws2hdf GGG.1636

The nws2hdf program will create the HDF file in the current directory with a name derived from the data file name. For the above example the HDF file name would be ‘GGG_1636.hdf’.

This page intentionally left blank.

Appendix H
GPS Sea Ice Motion Vectors
HDF-EOS User Documentation

The 'im2h' program reads GPS Sea Ice Motion Vectors data files distributed from the ASF DAAC and converts them to HDF formatted files. The inputs consist of a leader (metadata) file and a data file.

The program is written in ANSI C and requires the HDF library (4.0b1 or later) in order to compile. If you are using a Sun SPARC workstation with the HDF include libraries in the directory /usr/local/include and the library files (libdf.a and libnetcdf.a) in the directory /usr/local/lib, the following command will create an executable program im2h from the im2h.c source code file:

```
cc -DSUN -I/usr/local/include im2h.c -o im2h -L /usr/local/lib -lnetcdf -ldf -ljpeg
```

When using a machine other than a Sun, replace the -DSUN with the appropriate machine type in the above command line. For example, if you are using a Silicon Graphics machine, replace the -DSUN with -DIRIS4. A list of the supported machine types appears in the file HDF4.0BX/README where 'X' is the current release number.

To run the im2h program, type the program name followed by the leader (metadata) file name and the data file name.

Example: im2h 17766010.ldr 17766010.dat

The im2h program will create the HDF file in the current directory with a name derived from the data file name. For the above example, the HDF file name would be ‘17766010.hdf’.

This page intentionally left blank.

Appendix I
TOPEX/Poseidon Merged Geophysical Data Record HDF-EOS User Documentation

Every CD-ROM distributed by the Jet Propulsion Laboratory for the TOPEX/Poseidon Merged Geophysical Data Record (MGDR) data, consists of 2 ‘cycles’ of data. Each cycle contains approximately 254 ‘pass files’. A single TOPEX/Poseidon MGDR granule is a single ‘pass file’ of one ‘cycle’. The 't2h' program reads a single pass file and then converts it to an HDF formatted file. The input consists of a single file, which contains both the metadata and the data objects.

The program has been written in ANSI C and requires the HDF library (4.0b1 or later) in order to compile. If you are using a Sun SPARC workstation with the HDF include libraries in the directory /usr/local/include and the library files (libdf.a and libnetcdf.a) in the directory /usr/local/lib, the following command will create an executable program t2h from the t2h.c and the tlib.c source code files:

```
cc -DSUN  -I/usr/local/include t2h.c tlib.c -o t2h -L /usr/local/lib -lnetcdf -ldf -ljpeg
```

When using a machine other than a Sun, replace the -DSUN with the appropriate machine type in the above command line. For example, if you are using a Silicon Graphics machine, replace the -DSUN with -DIRIS4. A list of the supported machine types appears in the file HDF4.0bX/README where 'X' is the current release number.

To run the t2h program, type the program name followed by the data file name.

Example: t2h MGA12803.001

The t2h program will create the HDF file in the current directory with a name derived from the data file name. For the above example the HDF file name would be ‘MGA12803_001.hdf’.

This page intentionally left blank.

Appendix J
UARS SOLSTICE
HDF-EOS User Documentation

The 's2h' program reads UARS SOLSTICE level 3BS data files distributed from the GSFC DAAC and converts them to HDF formatted files. The UARS SOLSTICE granules consist of a metadata file (*META extension) and a data file (*PROD extension).

The program is written in ANSI C and requires the HDF library (3.3r4 or later) in order to compile. If you are using a Sun SPARC workstation with the HDF include libraries in the directory /usr/local/include and the library files (libdf.a and libnetcdf.a) in the directory /usr/local/lib, the following command will create an executable program s2h from the s2h.c source code file:

```
cc -DSUN -I/usr/local/include s2h.c -o s2h -L /usr/local/lib -lnetcdf -ldf
```

When using a machine other than a Sun, replace the -DSUN with the appropriate machine type in the above command line. For example, if you are using a Silicon Graphics machine, replace the -DSUN with -DIRIS4. A list of the supported machine types appears in the file HDF3.3rX/README where 'X' is the current release number.

To run the s2h program, type the program name followed by the metadata file name. It is assumed that the data file is in the same directory as the metadata file.

Example: s2h SOLSTICE_L3BS_D0022.V0007_C01_META

The s2h program will create the HDF file in the current directory with a name derived from the data file name. For the above example the HDF file name would be ‘SOLSTICE_L3BS_D0022.V0007_C01.hdf’.

This page intentionally left blank.

Acronyms

API	Application Program Interface
ASF	Alaska SAR Facility
AVHRR	Advanced Very High Resolution Radiometer
CDF	Common Data Format
CEOS	Committee on Earth Observing Systems
COTS	Commercial off the Shelf
DAAC	Distributed Active Archive Center
DIF	Directory Interchange Format
DMP	Data Migration Plan
ECS	EOSDIS Core System
EDC	EROS Data Center
EDHS	ECS Data Handling System
EOS	Earth Observing System
EOSDIS	Earth Observing System Data Information System
ERS-1	European Remote Sensing satellite - 1
ESA	European Space Agency
ESDISP	Earth Science Data Information System Project
FIRE	First ISCCP Regional Experiment
FTP	File Transfer Protocol
GB	Gigabytes
GCMD	Global Change Master Directory
GDR	Geophysical Data Record
GIF	Graphical Interchange Format
GPS	Geophysical Processing System
GSFC	Goddard Space Flight Center
HAIS	Hughes Applied Information Systems
HDF	Hierarchical Data Format

HDF-EOS	Hierarchical Data Format-Earth Observing System
HITC	Hughes Information Technology Company
HSTX	Hughes STX Corp.
IDL	Interactive Data Language
IMS	Information Management System
ISCCP	International Satellite Cloud Climatology Project
JPL	Jet Propulsion Laboratory
KB	Kilobytes
LaRC	Langley Research Center
MB	Megabytes
MGDR	Merged Geophysical Data Record
NASA	National Aeronautics and Space Administration
NCSA	National Center for Supercomputing Applications
NDVI	Normalized Difference Vegetation Index
netCDF	network Common Data Format
NGDC	National Geophysical Data Center
NOAA	National Oceanic and Atmospheric Administration
NSIDC	National Snow and Ice Data Center
NWS	National Weather Service
POES	Polar Orbiting Environmental Satellite
RIS	Raster Image Set
SAR	Synthetic Aperture Radar
SFDU	Standard Formatted Data Unit
SIS	Software Interface Specification
SDP	Science Data Plan
SSPP	Solar Stellar Pointing Platform
TOPEX	Ocean Topography Experiment
UARS	Upper Atmosphere Research Satellite
UDS	User and Data Services
WWW	World Wide Web

References

Version 1 Data Migration Plan (DMP), January 1995
URL: http://edhs1.gsfc.nasa.gov/

Proposed ECS Core Metadata Standard Release 2.0, December 1994
URL: http://edhs1.gsfc.nasa.gov/

Draft HDF-EOS Primer for Version 1 EOSDIS, January 1995
URL: http://edhs1.gsfc.nasa.gov/

Science Data Plan (SDP), July 1994, Version 3

ECS Core Metadata Migration Analysis, April 1995

Getting Started with HDF, Draft, Version 3.2, May 1993
URL: file://ftp.ncsa.uiuc.edu

HDF Reference Manual Version 3.3, February 1994
URL: file://ftp.ncsa.uiuc.edu

NCSA HDF Specification and Developer's Guide Version 3.2, September 1993
URL: file://ftp.ncsa.uiuc.edu